The Handbook of Evolutionary Economic Geography

Edited by Ron Boschma and Ron Martin

THE HANDBOOK OF EVOLUTIONARY ECONOMIC GEOGRAPHY

Dedication

Whilst this book was in press, we received the sad news that Bent Dalum passed away on 29 January, 2010. From the beginning, Bent was a highly enthusiastic supporter of the project of evolutionary economic geography, and a very positive contributor to this Handbook. We are extremely grateful for his wisdom and insights over the years. Above all, we will remember Bent as a most sympathetic and warm friend and colleague. We therefore dedicate this book to his memory.

The Handbook of Evolutionary Economic Geography

Edited by

Ron Boschma

Department of Economic Geography, Urban and Regional research centre Utrecht (URU), Utrecht University, The Netherlands

Ron Martin

Department of Geography and St Catharine's College, University of Cambridge, UK

Edward Elgar

Cheltenham, UK • Northampton, MA, USA

Published by
Edward Elgar Publishing Limited
The Lypiatts
15 Lansdown Road
Cheltenham
Glos GL50 2JA
UK

Edward Elgar Publishing, Inc.
William Pratt House
9 Dewey Court
Northampton
Massachusetts 01060
USA

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

Library of Congress Control Number: 2009941013

FSC
Mixed Sources
Product group from well-managed
forests and other controlled sources
www.fsc.org Cert no. SA-COC-1565
© 1996 Forest Stewardship Council

ISBN 978 1 84720 491 2 (cased)

Printed and bound by MPG Books Group, UK

Contents

PART 3 NETWORK EVOLUTION AND GEOGRAPHY

PART 4 INSTITUTIONS, CO-EVOLUTION AND ECONOMIC GEOGRAPHY

PART 5 STRUCTURAL CHANGE, AGGLOMERATION EXTERNALITIES AND REGIONAL BRANCHING

Contributors

Cristiano Antonelli Dipartimento di Economia “S. Cognetti de Martiis” Università di Torino and BRICK (Bureau of Research on Innovation, Complexity and Knowledge) Collegio Carlo Alberto, Moncalieri (Torino), Italy.

Ron Boschma Department of Economic Geography, Urban and Regional research centre Utrecht (URU), Faculty of Geosciences, Utrecht University, The Netherlands.

Giulio Bottazzi LEM and CAFED, Scuola Superiore Sant’Anna, Pisa, Italy.

Stefano Breschi KITeS, Università L. Bocconi, Milan, Italy.

Uwe Cantner Department of Economics, Friedrich-Schiller-University Jena, Germany.

Gabriele Cioccarelli Faculty of Economics, University of Pavia, Italy.

Philip Cooke Centre for Advanced Studies, Cardiff University, UK.

Michael S. Dahl Department of Business Studies, Aalborg University, Denmark.

Bent Dalum Department of Business Studies, Aalborg University, Denmark.

Carla de Laurentis Cardiff School of City and Regional Planning, Centre for Advanced Studies in Social Sciences, Cardiff University, UK.

Stefano Denicolai Faculty of Economics, University of Pavia, Italy.

Pietro Dindo LEM and CAFED, Scuola Superiore Sant’Anna, Pisa, Italy.

Jürgen Essletzbichler Department of Geography, University College London, UK.

Lee Fleming Harvard Business School, Boston, MA, USA.

Koen Frenken Eindhoven Centre for Innovation Studies, School of Innovation Sciences, Eindhoven University of Technology, and Urban and Regional Research Centre Utrecht, Faculty of Geosciences, Utrecht University, The Netherlands.

Elisa Giuliani DEA, Facoltà di Economia, University of Pisa, Italy, and SPRU, The Freeman Centre, University of Sussex, Brighton, UK.

Johannes Glückler Department of Geography, Heidelberg University, Germany.

Holger Graf Department of Economics, Friedrich-Schiller-University Jena, Germany.

Robert Hassink Department of Geography, University of Kiel, Germany.

Simona Iammarino Department of Geography and Environment, London School of Economics and Political Science, London, and SPRU, University of Sussex, Brighton, UK.

Jan Lambooy Department of Economic Geography, Faculty of Geosciences, Utrecht University, The Netherlands.

Camilla Lenzi KITeS, Università L. Bocconi, Milan, Italy.

Francesco Lissoni KITeS, Università L. Bocconi, Milan, and Department of Mechanical Engineering, Università di Brescia, Italy.

Anders Malmberg Department of Social and Economic Geography, Uppsala University, Sweden.

Ron Martin Department of Geography, and St Catharine's College, University of Cambridge, UK.

Peter Maskell Copenhagen Business School, Department of Innovation and Organizational Economics, Frederiksberg, Denmark.

Philip McCann Department of Economic Geography, Faculty of Spatial Sciences, University of Groningen, The Netherlands, and The University of Waikato, Hamilton, New Zealand.

Christian R. Østergaard Department of Business Studies, Aalborg University, Denmark.

David L. Rigby Department of Geography, University of California Los Angeles, CA, USA.

Jan W. Rivkin Harvard Business School, Boston, MA, USA.

Eike W. Schamp Department of Human Geography, University of Frankfurt, Germany.

James Simmie Department of Planning, Oxford Brookes University, Oxford, UK.

Olav Sorenson Yale School of Management, New Haven, CT, USA.

Udo Staber Department of Management, College of Business and Economics, University of Canterbury, Christchurch, New Zealand.

Erik Stam Utrecht School of Economics, Utrecht University, The Netherlands, Max Planck Institute of Economics – Entrepreneurship, Growth and Public Policy Group, Jena, Germany, University of Cambridge, UK, and Scientific Council for Government Policy, The Hague, The Netherlands.

Simone Strambach Faculty of Geography, Philipps University of Marburg, Germany.

Peter Sunley Department of Geography, University of Southampton, UK.

Andrea Vezzulli KITeS, Università L. Bocconi, Milan, Italy.

Antonella Zucchella Faculty of Economics, University of Pavia, Italy.

Acknowledgements

The idea for this book first emerged from an exploratory workshop on 'Evolutionary Economics and the Evolution of the Economic Landscape' organised by the editors at St Catharine's College, University of Cambridge, UK in April 2006. The workshop brought together nearly 30 economic geographers and economists from across Europe, with interests and expertise in evolutionary economics, to discuss the scope for and challenges of constructing a new evolutionary paradigm in economic geography. We are very grateful to the European Science Foundation (ESF Award: EW05-253 SCSS) for funding this exciting and highly fruitful event. A number of papers developed from the presentations at the workshop. Some of these were subsequently published in a special issue of the *Journal of Economic Geography*, in 2007. But such was the full range of papers that it was decided to bring them all together – and to solicit new versions of those that had already been published – into a major statement in the form of *The Handbook of Evolutionary Economic Geography*. Almost all of the contributions, therefore, are new to the Handbook, the exceptions being Chapter 15 by Sorenson, Rivkin and Fleming, which is reprinted – though with a new introduction – from *Research Policy*, **35** (7) (2006), and Chapter 4 by Martin and Sunley which is reprinted from the *Journal of Economic Geography*, **7** (5) (2007). We are grateful to Elsevier BV and Oxford Journals (Oxford University Press) respectively for permission to reprint these papers. Throughout the compilation of the book we have had been fortunate to have had the support and encouragement of Matthew Pitman of Edward Elgar. His encouragement and patience have been critical to the completion of this project. And last, but by no means least, we are indebted to the various contributors to this book: from their involvement in the initial workshop and through the subsequent ongoing dialogue around their chapters, their enthusiasm has been unwavering.

Ron Boschma and Ron Martin
August 2009

INTRODUCTION

THE NEW PARADIGM OF EVOLUTIONARY ECONOMIC GEOGRAPHY

1 The aims and scope of evolutionary economic geography

Ron Boschma and Ron Martin

1. Introduction

Over the past two and half decades, key theoretical developments have been taking place in the field of economic geography (Martin, 2008; Martin and Sunley, 2007b). For their part, economic geographers have moved firmly away from traditional economic analysis, and sought insights from various forms of heterodox economics and from other social sciences outside the economics field (see Amin and Thrift, 2000; Bagchi-Sen and Lawton Smith, 2006; Bathelt and Glückler, 2003; McCann, 2007; Martin and Sunley, 2001; Martin and Sunley, 2007a; Sheppard and Barnes, 2000; Simmie, 2005). Their interest has been in the institutional, cultural and social foundations of regional and urban development: a so-called 'institutional' or 'cultural turn' has taken place. At the same time, since the early to mid-1990s, several economists, led by Paul Krugman, the Nobel Laureate, on the one hand, and by Michael Porter, the business economist, on the other, have discovered geography, and argued for the importance of a geographical perspective for understanding the dynamics and competitiveness of the economy: both have emphasised the process of spatial agglomeration of economic activity as a source of increasing returns. Krugman and his followers even labelled their formal mathematical approach as the 'New Economic Geography'.

However, what has been lacking from these theoretical developments is any real appreciation of the importance of history in the economic landscape: neither perspective really tells us much about how that landscape *evolves over time*. Yet an evolutionary perspective is essential to a fuller understanding of such issues as the geographies of technological progress, dynamic competitive advantage, economic restructuring, and economic growth. In this context, there is thus considerable scope and potential for applying and extending the ideas and concepts from evolutionary economics to our analysis of regional and urban development.

Until the past few years, evolutionary economics, which itself has only developed in earnest since the early 1980s, has not attracted much attention from either economic geographers or the new breed of geographical economists. But very recently, a new evolutionary geographic perspective on the economic landscape has begun to emerge among geography and economics scholars, especially across Europe. This new body of work, though hitherto somewhat scattered, has gained sufficient momentum to warrant bringing the key conceptual, theoretical and empirical advances together in a clear statement on the aims, objectives and methods of this new paradigm. This is the central aim of this book. It is based on a special European Science Foundation Workshop on Evolutionary Economic Geography, held at St Catharine's College in the University of Cambridge, UK in 2006, which drew together a number of the most distinguished scholars in the fields of evolutionary economics and economic geography.[1] A basic conclusion of the workshop was that Evolutionary Economic Geography constitutes a distinctive and promising paradigm, and that the time is ripe for a major collective statement on the subject. While evolutionary economic geography has attracted increasing attention (and debate) since that workshop (e.g. Boschma and Martin, 2007, *Journal of Economic Geography*, 2007; Frenken, 2007; *Economic Geography*, 2009), there are as yet few such comprehensive statements.[2] This book should therefore play a formative role in influencing the future research agenda in this area.

The book covers both theoretical and empirical aspects of evolutionary economic geography. The contributions are grouped into five parts. The first set of chapters address some key theoretical and conceptual issues in evolutionary economic geography, with a dual focus not just on how ideas and concepts from evolutionary economics can be brought to bear on economic-geographic issues and settings, but also on how a geographical perspective itself has implications for our notions of economic evolution. Against this conceptual background, the rest of the book is concerned much more with applying these ideas in specific and empirical contexts. But here too, in so doing the various chapters also make important contributions to the formulation of an evolutionary perspective on the economic landscape. In the second part of the book, the focus shifts from broad conceptual issues to the specific case of firm and industrial dynamics in space. The contributions in this section explore how an evolutionary framework can be fruitfully applied to topics in economic geography at both the micro level (like the geography of entrepreneurship) and the meso level (like the tendency of industries to cluster in space). Since networks play a crucial role in understanding the spatial uneven distribution of economic activity, the third part of the book is devoted to the nature and spatial evolution of networks. Part 3 concentrates on how networks may be integrated in evolutionary economic geography. The fourth part focuses on the evolution of institutions in territorial contexts and explores how institutions may be incorporated in the explanatory framework of evolutionary economic geography. Part 5 of the book deals with the evolution of agglomerations and the economic landscape from an evolutionary perspective. This introductory chapter outlines the main issues addressed in each section of the book, and identifies the main arguments of the various contributions.

2. The aims and conceptual foundations of evolutionary economic geography

Constructing an evolutionary economic geography, though an exciting endeavour, is by no means a straightforward task. For one thing there is not a single, generally accepted and commonly used body of evolutionary economics to draw on for inspiration. To be sure, over the past two decades or so a new evolutionary economics has rapidly emerged that seeks to understand precisely how the real economy evolves through real time (see, for example, Arthur et al., 1997; Dopfer, 2004; Foster, 1997; Hodgson, 1993; Metcalfe, 1998; Metcalfe and Foster, 2004; Nelson and Winter, 1982; Potts, 2000; Witt, 2003, 2006). But the rush of enthusiasm to adopt an 'evolutionary perspective' has tended to produce a plethora of self-declared approaches – a 'massive hybridisation of theory', as Dopfer and Potts (2004, p. 195) put it – rather than a single coherent body of concepts and methods. So to some extent, economic geographers face a still developing corpus of ideas. Nevertheless, although the field of evolutionary economics is still without 'stabilised shared meaning' (Klaes, 2004), and remains somewhat embryonic, some basic principles do seem to be crystallising.

According to Witt (2003, 2006), the key focus of evolutionary economics is on the processes and mechanisms by which the economy *self-transforms itself from within*. Thus theories on economic evolution have to satisfy three basic requirements. First, they must be *dynamical*. This criterion rules out any kind of static or comparative-static analysis, and focuses attention on change. Second, evolutionary economics must deal with *irreversible* processes – the past cannot be recovered and it imparts legacies that condition the behaviour of economic agents in the present and the future; this rules out all 'dynamical' theories that describe stationary states or equilibrium movements, the preoccupation of traditional mainstream (neoclassical) economics. Rather, in the context of evolutionary economics, 'dynamical' refers to such features as emergence, convergence, divergence, and other patterns and trajectories that are rooted in real historical time. This distinction is critical since while mainstream economists – and the 'new economic geography' theorists – claim to deal with 'history', this notion is merely a logical construct relating to the 'initial conditions' of the abstract mathematical models used to determine stable equilibrium outcomes: there is no real history in such approaches. And third, theories on economic evolution must cover the generation and impact of *novelty* as the ultimate source of self-transformation. As Witt emphasises, the criterion of novelty – its generation and its role in economic transformation – is crucial to any theory of economic evolution. It is the creative capacity of economic agents (individuals and firms), and the creative functions of markets, that drive economic evolution and adaptation (see also Metcalfe et al., 2006).

As Schumpeter insisted, transformation arises endogenously, from within the socio-economic system, and enterprise-driven innovation and adaptive development are the primary processes (Ramlogan and Metcalfe, 2006). Thus innovation and knowledge assume central importance in evolutionary economics. Knowledge is not something that is separate from, or autonomous to, the economic process in the manner of some pregiven 'factor of production' (as it is in so-called 'endogenous growth' models); rather it is the internal development of knowledge that renders the underlying process of economic evolution both adaptive and transformative in character (Fine, 2000). Knowledge never stands still, but is constantly being created. It is this continual process that drives economic evolution, and renders capitalism restless, in constant motion:

> The origins of restless capitalism lie in its unlimited capacity to generate knowledge and new behaviour from within, and it is the propensity for endogenous variation that makes it so dynamic and versatile, sufficiently so that economies may be completely transformed in structure over relatively short periods of historical time. Growth is not simply a result of calculation within known circumstances, but of human imagination and the search for novelty and competitive advantage. Moreover, every advance in knowledge creates the conditions for further advances . . . economic growth is an autocatalytic process in which change begets change. (Metcalfe et al., 2006, p. 9)

Thinking about the economy as a dynamical, irreversible and self-transformational system opens up new space for theoretical, ontological and epistemological exploration. Indeed, as noted by Dopfer, Potts, Klaes and Witt among others, the attraction for many of evolutionary economics is precisely its permissiveness towards heterodox perspectives and approaches, and a range of different metaphors (see also Castellacci, 2006). Thus it is possible to identify neo-Veblenian, neo-Schumpeterian, neo-Hayekian, and neo-Darwinian approaches. For many, the challenge is to give economic interpretation to the basic ideas of modern evolutionary biology – especially to the notions of variety, selection, fitness, retention, mutation and adaptation. For others, the notions from complexity science – such as self-organisation, co-evolution, emergence, far-from equilibrium dynamics, and criticality – provide a suitable conceptual framework. For yet others, it is the combination of these two perspectives that is most promising.

The upshot is that evolutionary economics offers a rich palette of ideas and concepts for geographers to draw on to help them explain the evolution of the economic landscape. Here of course is the second difficulty. As is the case with any school of economic thought, evolutionary economics is largely aspatial in outlook and formulation, whereas geographers are interested in applying and adapting concepts from evolutionary economics to spatial contexts and processes. As numerous authors have warned, the abduction of metaphors and concepts from one field into another can be problematic (see Wimmer and Kössler, 2006). Just as the use of physical and mechanical analogies and metaphors in mainstream economics is contentious, so the use of biological analogies and concepts in evolutionary economics – and evolutionary economic geography – is not uncontroversial. How far and in what ways the notions taken from traditional and modern evolutionary biology can be translated into meaningful economic equivalents is itself a topic of lively discussion and debate (see, for example, Foster, 2000). Applying a paradigm from one science to another is a risky venture. For example, using biological analogies blindly or slavishly, without due care for inappropriate ontological transfers, will of course hardly constitute a theory of economic evolution (see Mokyr, 2005). However, according to Mokyr, Metcalfe and others, the argument is not so much that the economy is in some ways 'similar' to biological systems, but that Darwinian models and related concepts (of selection, variety, novelty, etc.) transcend biology, and that, indeed, evolutionary biology is just a special case of a much wider and broader set of models that try to explain how certain kinds of system evolve over time. Thus, despite the risks and inherent dangers involved, importing metaphors, concepts and methodologies from other disciplinary fields remains one of the major sources of theoretical and empirical innovation, not only providing new perspectives but also in the process stimulating conceptual advance and creating new intellectual contact points and avenues for cross-disciplinary cooperation.

Such potential benefits are undoubtedly a major factor stimulating evolutionary approaches to economic geography. However, it is not simply a case of applying such concepts and their theoretical and methodological frameworks to economic geography, though this in itself is a challenging enough task. For an evolutionary economic geography cannot simply be derivative in its ambitions. The goal is twofold: not only to utilise the concepts and ideas from evolutionary economics (and evolutionary thinking more broadly) to help interpret and explain how the economic landscape changes over historical time, but also to reveal how situating the economy in space adds to our understanding of the processes that drive economic evolution, that is to say, to demonstrate how *geography matters in determining the nature and trajectory of evolution of the economic system*. The contributions to this book are all motivated by this dual ambition.

What then are the aims, the distinguishing features, of an evolutionary approach to economic geography? Put broadly, we can say that the basic concern of evolutionary economic geography is with *the processes by which the economic landscape – the spatial organisation of economic production, circulation, exchange, distribution and consumption –*

is transformed from within over time. As Boschma and Martin (2007) put it, evolutionary economic geography is concerned with the spatialities of economic novelty (innovations, new firms, new industries, new networks), with how the spatial structures of the economy emerge from the micro-behaviours of economic agents (individuals, firms, organisations); with how, in the absence of central coordination or direction, the economic landscape exhibits self-organisation; and with how the processes of path creation and path dependence interact to shape geographies of economic development and transformation, and why and how such processes may themselves be place dependent. Our concern is both with the ways in which the forces leading to economic change, adaptation and novelty shape and reshape the geographies of wealth creation, work and welfare, *and* with how the spatial structures and features so produced themselves feed back to influence the forces driving economic evolution. For the economic landscape is not just the passive outcome or by-product of the process of economic evolution, but a conditioning influence on that process. Economic transformation proceeds differently in different places, and the mechanisms involved neither originate nor operate evenly across space. The emphasis is on understanding the processes and mechanisms that make for or hinder the adaptation of the economic landscape, and how spatial and historical contingency interact with systemic necessity.

Given these aims, what are the theoretical and conceptual foundations on which such an understanding might be based? This question is the focus of Part 1 of the book, where several authors explore and assess some of the possible conceptual foundations for an evolutionary economic geography. Within economics – and indeed in other social sciences – it is possible to identify three main approaches to the study of evolution: generalised Darwinism, the theory of complex adaptive systems, and path dependence ideas (see Figure 1.1).[3] Much of evolutionary economics is based on ideas and concepts taken

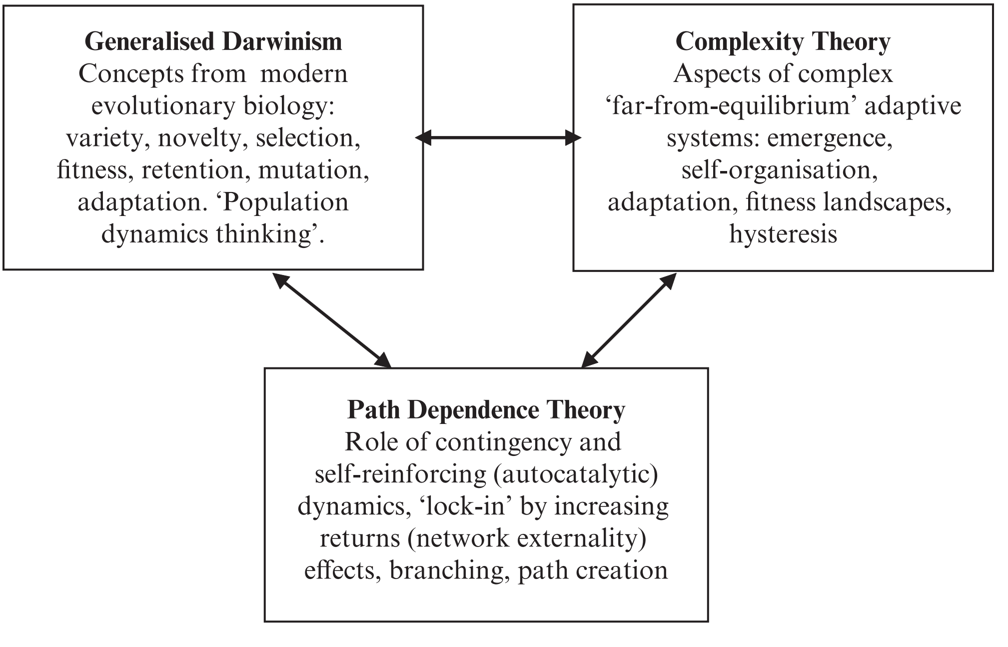


Figure 1.1 Three major theoretical frameworks for evolutionary economic geography

from generalised Darwinism, especially those of variety, selection, novelty and retention (see, for example, Metcalfe, 2005; Witt, 2003). By comparison, complexity theoretic ideas have received less attention, although the potential of this approach is increasingly recognised, with some authors linking complexity concepts explicitly with the analysis of economic evolution (e.g. Beinhocker, 2006; Foster, 2005; Potts, 2000; Rosser, 2009). The third approach, based on path dependence, and based especially on the writings of Paul David and Brian Arthur, is concerned with giving economics a prominent historical dimension, and has been a key ingredient of many versions of evolutionary economics. Although distinctive frameworks, there are overlaps between the three approaches, and hybrid frameworks that combine elements from two or all three.

Likewise, in evolutionary economic geography it has been the generalised Darwinism perspective that has been most frequently invoked, closely followed by the path dependency approach: by comparison, complexity theoretic notions have yet to be explored in any concerted way. The basic contours of an evolutionary model of economic dynamics based on the principles of generalised Darwinism are outlined by Essletzbichler and Rigby in Chapter 2. Employing the core evolutionary principles of variety, selection and retention (continuity), they argue that competition between agents located in different geographical spaces may produce distinct economic regions. While certainly not being units of selection, regions can be conceptualised as *selection environments* within which, and across which, evolutionary processes operate. The authors claim that evolutionary economic geography should focus on the evolution of a population of agents within a single region, as well as on the evolution of different regions that might, or might not, affect the dynamics of each other's populations. These arguments are extended to illustrate how emergent properties of economic agents and places co-evolve and lead to different trajectories of economic development over space. While there remain numerous issues to be resolved and elaborated, as these authors demonstrate, an approach based on notions and principles derived from generalised Darwinism offers economic geographers a theoretically rich framework for the analysis of change within the economic landscape. It certainly is a framework that informs many of the chapters in this handbook.

Perhaps the most often used notion used in economic geographic work that has sought to take history seriously in studies of regional development is that of path dependence, that is the idea that the economic landscape does not tend towards some (predefined) unique equilibrium state or configuration, but is an open system that evolves in ways shaped by its past development paths. As Martin and Sunley (2006) argued in their extensive review of the notion, the idea of path dependence has been taken as a fundamental principle of economic evolution by numerous economic geographers. But as Martin and Sunley argued in that work, and even more forcefully in Chapter 3, this assumption is by no means unproblematic. For one thing, there is the problem of defining what it is about regional economies that follows a path dependent trajectory of development – the region's firms, its industries or the regional economy as a whole? Can multiple paths co-exist, and how do they interact? Second, what are the processes that allegedly engender path dependence in the economic landscape? Further, where do new paths come from, and why do they emerge where they do? And how do old paths come to an end? These questions have not received the critical attention they require. Martin and Sunley's (2006) extensive discussion sought to stimulate just that sort of discussion. Here, however, they are explicitly concerned to elucidate the question of what sort of evolution is implied by the concept of path dependence. They take issue with Paul David's (2005) recent talk of 'path dependent equilibrium economics' on the grounds that the notion of equilibrium is antithetical to that of evolution. The issue is not resolved, in their view, by David's interpretation of path dependence as the historically contingent selection of, and eventual lock-in to, one of a number of possible multiple equilibria outcomes or states. They argue that the case of a technology, industry or regional economy becoming locked into one of a number of possible multiple equilibrium states may be the exception rather than the rule, and that we therefore need a more open notion of path dependence that allows for more or less continuous adaptation and mutation of technologies, industries and regional economies (see also Martin, 2010b). Such adaptation and mutation is almost certain to be path dependent in nature, and suggests that development paths need not ever reach any kind of equilibrium, that such paths can atrophy and decline over time as a result of endogenous processes (and not because of 'external shocks', as assumed by David-type path dependence models), and that new paths can emerge out of existing ones.

Compared to the use of the generalised Darwinian and path dependence approaches, much less attention has been directed to constructing an evolutionary economic geography based on complexity theory. According to Foster (2005), evolutionary biology, with its focus on selection mechanisms is limited in its applicability to socioeconomic contexts. For this reason he advocates the exploration of a more morphogenetic perspective that draws on complexity-theoretic notions. Although this approach has it roots in a more fundamental physical level of enquiry, namely non-equilibrium thermodynamics, Foster argues that it offers a more useful analytical representation of structuration for those interested in detecting evolutionary change in time series data. As such, Foster contends, the complexity approach is not used as a metaphor or analogy – as is the case with the generalised Darwinian perspective – because structuration processes are present at all levels of scientific enquiry, including the socioeconomic.

In Chapter 4, Martin and Sunley take up this line of enquiry to explore the idea of the economic landscape as a complex adaptive system. They identify several key notions of complexity theory and what is being called the new 'complexity economics' (as championed recently by Beinhocker, 2006, see also Rosser, 2009), and examine whether and in what ways these ideas can be used to help inform an evolutionary perspective for understanding the uneven development and transformation of the economic landscape. Complexity theory deals with open systems subject to constant interaction with their environments, that are dynamic, typically 'far-from-equilibrium', yet which display internal order and the emergence of structure (self-organisation). As Martin and Sunley point out, these notions resonate closely with questions about how the spatial structure of an economy emerges and changes; about how regional and urban economies rise and fall in relative prosperity; about why some regional and urban economies appear more adaptable and resilient than others over time to shifts in technology, markets, policy regimes and the like; about why certain industries and technologies develop in particular geographical areas but not others; and about how the various spatial networks of economic relationships and flows form and evolve. In this sense they argue that complexity thinking could make a valuable contribution to the construction of an evolutionary economic geography. But they also express reservations over the increasingly dominant modelling paradigm associated with complexity analysis in economics, including the functional development of appropriate computational architectures (such as multi-agent models and dynamical systems models), and instead urge a more philosophically inclined social-ontological approach. What precisely does it mean to talk of the economic landscape as a complex system? In what sense is the economic landscape a meaningful complex system to which the concepts of complexity thinking can be meaningfully applied? What does connectivity mean and how do we distinguish partial from strong connections? These are difficult questions. To be analytically useful, complexity is not something that just bolts on to or can be blended with an existing conceptual/theoretical framework to add a 'complexity perspective' or 'evolutionary perspective'. Nor is it sufficient to invoke the terminology and concepts of complexity science without thinking through what these concepts are being applied to, and what they mean in an economic-geographical context. Challenging questions though these are, in the view of Martin and Sunley the answers could well be rewarding.

A way of dealing with complex systems from an evolutionary perspective is to analyse how networks of agents evolve over time. While the study on network evolution is still in a premature phase (see for example Powell et al., 2005), there is growing interest from researchers to employ social network tools to describe and explain the evolution of network structures and their performance over time. Economic geographers have started to contribute to this emerging body of literature only quite recently. They are increasingly aware that knowledge networks, and their spatial configuration, play a crucial role in the innovation process, and therefore may be considered a driving force of the evolution of the economic landscape. However, network analysis is still very much underdeveloped in the geography of innovation, and this is certainly true for an evolutionary approach to this topic, though work is beginning to emerge (see, for example, Giuliani, 2007; Glückler, 2007; and some of contributions to this handbook, such as Breschi et al., Cantner and Graf, Giuliani, and Glückler; Chapters 16, 17, 12 and 14 respectively).

In Chapter 5, Boschma and Frenken take up this challenge by proposing an evolutionary perspective on the spatial evolution of innovation networks. They draw on insights from the proximity literature (Boschma, 2005; Torre and Rallet, 2005) to explain the evolution of the structure and performance of networks. Boschma and Frenken conceive different forms of proximity as alternative driving forces of network formation, geographical proximity being one of them. They propose an evolutionary perspective on the geography of network formation that is firmly embedded in a proximity framework (see also Sorenson et al., Chapter 15). Boschma and Frenken claim that proximity may be considered a prerequisite for agents to connect with and to enhance knowledge spillovers, but proximity between agents does not necessarily increase their innovative performance, and may possibly even hinder it. The authors refer to this as the 'proximity paradox'. Boschma and Frenken then turn to the long-term dynamics of networks, and discuss how such dynamics may be related to the changing role of proximity in the formation and performance of innovation networks. In this respect, crucial questions are how far and in what ways different proximities induce path dependence in the spatial evolution of networks (see, for example, Glückler, 2007; Ter Wal, 2009), and how this process depends on spatial context. This is an area of research that is still strongly underdeveloped, though considered crucial for the further development of an evolutionary perspective on the spatial evolution of networks. According to Boschma and Frenken, the ultimate goal is to develop a dynamic network approach that also accounts for the fact that the spatial evolution of network structures may, in turn, affect the degree of the different forms of proximity. That would really contribute to our understanding of the spatial evolution of networks as a truly endogenous process.

3. Firm dynamics, industrial dynamics and spatial clustering

In Part 2 of the book, the focus shifts from broad conceptual issues to the specific case of firm dynamics and industrial dynamics in space. Here, we start from the micro level, focusing on the locational behaviour of firms, and how firms compete and learn on the basis of their routines in time and space. This leads to firm dynamics in the economic landscape: new firms will enter the market, some firms will do well and increase their market share in an industry, while other firms will stagnate and exit the market. Moving to the meso level, one can investigate how these firm dynamics lead industries to evolve through different stages of development in time and space. The contributions in Part 2 of the book explore how such an evolutionary framework can be fruitfully applied to topics in economic geography at the micro level of firms (like the geography of entrepreneurship) and the meso level of industries (like spatial clustering). More particularly, the contributions examine how firms behave, compete and learn in space, whether firm dynamics at the industry level lead to spatial clustering, and whether spatial clustering brings positive or negative externalities to cluster firms along the life-cycle of an industry.

According to Boschma and Frenken (2006), evolutionary economic geography examines how the spatial structure of the economy emerges from the micro-behaviour of individuals and firms.[4] Instead of describing the behaviour of individuals and firms as if they optimise, they follow Simon's (1955) concept of bounded rationality to claim that firms are subject to cognitive constraints (Dosi et al., 1988; Nelson and Winter, 1982). In order to reduce uncertainty, firm behaviour is guided and constrained by routines. Because of their tacit and cumulative nature, routines are not easy to change, and are very difficult to imitate for other firms (Heiner, 1983). Evolutionary theory predicts that most firms innovate incrementally, exploiting the knowledge they have built up in the past. Nelson and Winter (1982) have described this as a 'local search process'. And when firms diversify and grow, they tend to expand into related products, that is, into those products that are technologically related to their current products (Penrose, 1959).

Taking such a micro-perspective, an evolutionary approach to economic geography can describe the evolution of the economic landscape as changes in the time-space distribution of routines over time (Boschma and Frenken, 2003), that is, how new routines come into existence, and how they diffuse in time and space. The economic landscape (as it manifests itself in the spatial clustering of industries, for instance) is then the result of an evolutionary sequence in which some variations of routines have been selected because, for some reason, they are better adapted than others. As mentioned above, selection occurs at the micro level of the firm (through its routines), but also at the macro level of markets and institutions. Market competition acts on variety as a selection device, which determines which (old and new) routines survive and prosper, and which ones decline and go out of business (Ormerod, 2005). In a dynamic economy, fitter routines become more dominant over time through selection, enabling more efficient firms with fitter routines to expand their production capacity and market shares at the expense of less efficient firms. The selection environment not only includes markets but also institutions, whose effects become especially visible when a major institutional change occurs and the 'playing field' on which firms compete changes dramatically. Thus, understanding the fitness of routines requires an analysis not only of firms and markets but also of institutions as relevant enabling and constraining contexts.

More importantly, fitter routines also expand in an economy through learning and routine replication within and between firms. Crucial in an evolutionary account is the idea that replication and reproduction of routines are imperfect (Winter and Szulanski, 2001). The processes by which routines are replicated, diffused, and further replicated in successive populations of firms often work ineffectively, which explains the existence and persistence of variation. This is not, however, a random process: these routines spread or disappear in a context that is biased cognitively, socially – and also geographically. So some variations can be learned and passed on, though this depends on the extent to which agents are proximate in various dimensions (spatially, relationally or cognitively – see Boschma, 2005). As Cohen and Levinthal (1990) have argued, firms can understand, absorb and implement external knowledge only when it is close to their own knowledge base. In other words, effective knowledge transfer between firms requires absorptive capacity and cognitive proximity (Nooteboom, 2000). While evolutionary economics has focused almost exclusively on the cognitive dimension, one can think of other forms of proximity that affect the successful replication and diffusion of routines within and between firms across space. So the evolution of the economic landscape is mediated by cognitive, social, geographical and institutional processes that warrant attention in evolutionary economic geography. As Boschma and Frenken in Chapter 5 argue, this opens up a new research agenda in which the proximity framework becomes embedded in an evolutionary approach to economic geography.

Putting emphasis on selection forces and constraints does not necessarily mean that human agency does not play a role in evolutionary economic geography (see, for example, Boschma and Frenken, 2006; Staber, 1997). On the contrary, the role of entrepreneurs is crucial. How new routines are created by entrepreneurs in space, how they transform the economic landscape, and how the economic landscape itself impacts on the geography of entrepreneurship are the topics that are taken up by Erik Stam in Chapter 6. The study of the location of new firms from a micro-evolutionary perspective goes back to Pred (1967) who also made use of the concept of bounded rationality. There is overwhelming evidence that new firms do not opt for optimal locations in terms of cost-minimization. Instead, they are affected by local structures laid down in the past, like the place of residence of the entrepreneur (many firms start from the entrepreneur's home), the location of the parent firm (spinoffs typically locate in the vicinity of their parent), the social networks of the entrepreneur (which are very local, and provide access to resources), and the regional knowledge base (new firms typically exploit local knowledge and skills). Critically surveying the geography of entrepreneurship literature, Erik Stam notices that entrepreneurship not only tends to be a geographically localised phenomenon, but it is also a spatially uneven process that tends to persist over time. This implies regional entrepreneurship tends to be a path-dependent process (see also Martin and Sunley, Chapter 3). For example, as most entrepreneurs start their venture from home, it may be expected that the current spatial distribution of people will explain to a considerable degree regional entrepreneurship rates, and as many successful entrepreneurs originate from existing firms, the current distribution of firms is most likely to affect the geography of successful entrepreneurship to some extent. These questions, among others, are crucial to develop an evolutionary economic geography framework that is embedded in a dynamic view on entrepreneurship.

Besides the question of where entrepreneurs create new routines in space, an evolutionary approach in economic geography is also concerned with how routines develop further within firms, and in relation to their regional environment. In Chapter 7, Cristiano Antonelli addresses the importance of pecuniary knowledge externalities for localised learning and cluster development. While technological knowledge externalities have gained most attention and are often assumed to be 'in the air' (to use Marshall's phrase), Antonelli assigns pecuniary externalities a central role in the generation and exploitation of knowledge. He explains why technological learning is spatially biased, and firms will use idiosyncratic production factors that are locally abundant. This provides a theoretical explanation for the development of distinctive competences within firms, for the phenomenon of localised learning, and the spatial clustering of firms.

When an evolutionary approach in economic geography is used to deal with the spatial evolution of routines, interest focuses especially on how firms employ and develop spatial strategies over time (see for example Brouwer, 2005). Routines are diffused across the economic landscape through organisational practices like the relocation of firms (Knoben, 2007; Stam, 2007), merger and acquisition activity of firms located in different places, and the establishment of new plants by incumbent firms in other locations (Wintjens, 2001). Analysing the evolution of spatial strategies of firms, evolutionary economic geography has the potential to contribute to a better understanding of the process of globalisation through the spatial diffusion of routines within and between firms.

In Chapter 8, Simona Iammarino and Phil McCann provide an explanation for why the strategies of multi-national entreprises (MNEs) result in a pattern of 'concentrated dispersion' worldwide. They claim that firms accumulate different competences in time and space, which impacts on their incentives to co-locate and tap into complementary knowledge bases in different locations. This brings up a range of issues that are very relevant to an evolutionary approach in economic geography, for example: what kind of strategies do MNEs follow to overcome geographical and other barriers in order to access distant knowledge bases, and how can host regions successfully benefit from incoming MNEs in terms of knowledge transfer (see, for example, Cantwell and Santangelo, 2002; Morgan, 1997; Morrison, 2008)? And crucial for an evolutionary approach, what are the dynamics in the relationship between MNEs and the host regions (e.g. Cantwell and Iammarino, 2003). Research suggests that, in certain circumstances, the multinational firm may become more embedded in the host region over time, and will transform the local environment of the host region accordingly (see, for example, Dunning, 1994; Peng, 1995; Storper, 1997; Wintjens, 2001).

As Iammarino and McCann also point out in their contribution, an evolutionary approach provides a powerful framework to explain the tendency of economic activities to agglomerate and of industries to cluster spatially (see Malmberg and Maskell, 1997, 2002). Because routine replication and processes of learning are often subject to failure, it helps to be co-located. What is more, routine replication and knowledge accumulation tend to operate at the regional level because the mechanisms through which they operate (like spinoff activity, firm diversification, labour mobility and social networking) tend to have a regional bias (Boschma and Frenken, 2009b): spinoffs tend to locate near their parent firm, new divisions of firms are frequently created inside established plants in the same location, employees change jobs primarily within the same labour market area, and social networks through which knowledge flows tend to be geographically localised. According to Boschma and Frenken (2009b), this implies that the lineage structure between routines is spatially dependent: once particular routines become dominant in certain regions, the subsequent evolution of these routines into the same and related industries will occur primarily in the same region.

There is a long tradition in economic geography that explains the spatial clustering of an industry from a sector life-cycle approach (e.g. Chapman, 1991; Markusen, 1985; Norton, 1979; Scott, 1988; Scott and Storper, 1987; Storper and Walker, 1989). Only more recently, this topic has been studied more systematically by investigating the (spatial) dynamics of the whole population of firms at the industry level through its different life-cycle stages (Boschma and Frenken, 2003). These industry life-cycle approaches analyse the spatial evolution of an industry in terms of entry, growth and exit of firms over time (Arthur, 1994; Klepper, 1997; for a critical review, see Boschma, 2007). Since firms' relations at the sector level are mainly of a competitive nature, entry-and-exit models and survival analysis are techniques that are often employed. The core models on the spatial evolution of industry are the organisational ecology framework as developed by Hannan and colleagues (Carroll and Hannan, 2000; Hannan and Freeman, 1989; Hannan et al., 1995; Wenting and Frenken, 2008; Wezel, 2005) and Klepper's industry life-cycle model (Klepper, 2007). These approaches provide additional insights to the extensive but rather descriptive literature on clusters. Instead of taking a static view on clusters, they provide a dynamic view on how clusters emerge and develop. Especially Klepper's industry lifecycle model has attracted attention, because it explains clustering from spinoff dynamics. Doing so, it provides a new, alternative explanation for clustering as the outcome of a spinoff process through which routines are passed on from incumbents (parent organisations) to new entrants (offspring). The Klepper model shows that such a process may lead to spatial clustering even in the absence of agglomeration economies. Interesting as that may be, evolutionary economic geography is not only concerned with the question how evolutionary processes (such as spinoff dynamics) impact on the geography of industries, but also how the economic landscape impacts on these evolutionary processes (Boschma, 2007; Boschma and Frenken, 2003; Martin, 1999; Martin and Sunley, 2006).

There are a growing number of studies that explain the spatial clustering of an industry from such an evolutionary framework. In Chapter 9, Michael Dahl, Christian Østergaard and Bent Dalum provide a clear example. Their contribution departs from the conventional literature that often takes a static approach to clusters, and which tends to focus on explaining dynamics of clusters ex post, referring to location-specific advantages, such as knowledge spillovers, networks and labour market pooling (for a critique of that approach, see, for example, Boschma and Lambooy, 1999; Martin and Sunley, 2003; Storper and Walker, 1989). Instead, they argue there is a need to view possible advantages of clusters more as an outcome of, rather than a precondition for the formation and early growth of clusters. But even so, their case study shows that the wireless communication cluster in Northern Denmark was the outcome of the initial success of some pioneering firms that gave birth to successful spinoff companies in the region. This 'success breeds success' story is in line with findings in other studies (Boschma and

Wenting, 2007; Buenstorf and Klepper, 2009; Klepper, 2007), which show that spinoffs are often successful because they can exploit knowledge acquired in very successful parent organisations.

This literature on industrial dynamics focuses on the mechanisms that drive the evolution of a population of firm-specific routines in an industry, and which make fitter routines more dominant in an industry. A key issue concerns the mechanisms through which successful routines diffuse across firms and cluster spatially when a new industry emerges. As noted above, spinoff dynamics and agglomeration economies may act as vehicles through which knowledge and routines are created and diffused among a growing population of firms within a territory. As noted above, spinoff dynamics is considered a driving force behind the spatial formation of an industry, because it diffuses relevant knowledge from incumbents (parents) to new firms (spinoffs) at the local level. Moreover, agglomeration effects may become manifest, once spatial clustering occurs. For instance, local knowledge spillovers may become increasingly available, which then cause a further spatial concentration of the industry. As such, spinoff dynamics and agglomeration economies provide alternative explanations for the spatial clustering of an industry. However, these may also be complementary, since spinoff activity in a region may strengthen agglomeration forces, which, in turn, may enhance the creation and survival rate of spinoffs (Boschma and Frenken, 2003).

In a study of the long-run evolution of the British automobile industry, Boschma and Wenting (2007) found that both effects play a role, but at different stages of the industry life-cycle. Spinoff companies did not show a higher survival rate during the first stage of the life-cycle, because there is still not much to be learnt from a parent in the same industry, since a dominant design is still lacking. By contrast, new entrants with working experience in related industries did well at this stage. Spinoffs performed better only in a later stage of the industry life-cycle, because pre-entry working experience in the same industry then appeared to be of much higher value. The effect of the location on firm's survival also differed along the industry life-cycle. Start-ups in the British car industry that were founded in regions with related industries had a higher survival rate during the first stage (see also Buenstorf and Klepper, 2009). While one would expect a positive effect of localisation economies on a firm's survival, Boschma and Wenting (2007) found no effect in the first stage of the industry life-cycle, and even a negative effect at a later stage. The more spatially concentrated the automobile industry became, the harder it was for new entrants to survive in clusters, probably because of more local competition (see for similar results Otto and Kohler, 2008; Staber, 2001). This provides an alternative view on clusters from the more conventional one, in which clusters are supposed to deliver what they ought to do, that is, bring benefits to local firms (see Staber's contribution in Chapter 10). This is in line with recent studies that show that clusters not only stimulate entries (being attractors to new firms) but also enhance exits (clusters being tough selection environments at the same time) (Wenting, 2008).

This is not to deny that spatial clustering in certain industries may be attributed to traditional cost factors, as resource intensive industries locate in the vicinity of natural resources or near transport nodes, and many service sectors locate close to the market. However, for many industries, there may be no obvious location in which to cluster, as their main input is knowledge from people (Lambooy, 2002). This knowledge – for the most part – has to be created alongside the development of the industry itself. The location of these knowledge-intensive industries may then be understood from evolutionary processes through which the industry gradually emerged and developed. This is not, however, to deny that location-specific characteristics may matter (see, for example, Bathelt and Boggs, 2003; Boschma and Lambooy, 1999; Brenner, 2004; Glaeser, 2005; Martin and Sunley, 2006). On the contrary, the pre-existing localised presence of knowledge, skills and creativity may still be a good reason why knowledge-intensive industries cluster in certain areas, and not in others, as the example of the spatial clustering of the British automobile industry demonstrated. However, instead of saying that locations matter in a deterministic manner (as propagated by neoclassical thinking), an evolutionary approach to spatial clustering accounts for dynamic processes through which routines are created and diffused (e.g. through the spinoff process, the formation of a specialised local labour market, or the establishment of institutions), and pre-existing structures in regions are expected to condition, but not determine their spatial outcome. This emphasis on contingency reflects the idea that locations may influence spatial clustering of industries to some extent, and it is up to empirical research to determine whether and to what extent this locational influence varies from industry to industry (Boschma, 2007; Boschma and Frenken, 2003). In order to determine the effect of pre-existing regional structures, it is crucial to assess the extent to which individual features of firms (i.e. their routines) matter for their survival. While many cluster studies tend to assume that clusters have a positive effect (why did the industry cluster otherwise?), this can only be concluded when one controls for the effect of firm-specific features. This is crucial, as Erik Stam observes in Chapter 6: entrepreneurship studies often show that personal attributes of entrepreneurs (like age, experience) provide a better explanation of entrepreneurial success than do regional features (see also e.g. Sternberg and Arndt, 2001).

Adopting a focus on individuals and firms could also contribute to a better understanding of the notion of routines, and how learning shapes and modifies routines (Hodgson, 2009). This could be accomplished by looking at how organisational routines, being a collective property of a firm (Nelson and Winter, 1982), are affected by entrepreneurial activities ('intrapreneurship') and the recruitment of new employees. Udo Staber explores in Chapter 10 how knowledge may be transmitted between firms, and how that affects the routines of firms. Staber proposes that we regard ideas in the form of meanings as the basic unit of transmission. Ideas are not isolated phenomena, but evolve as bundles. An evolutionary account could focus on how these bundles of ideas remain intact (like the set of skills in routines), but also how new ideas (through labour mobility or spinoff activity) may cause tension and incoherence in the existing set of ideas. In a similar context, Wenting (2008) has made an attempt to assess the effect of post-entry labour mobility on firm performance in the fashion industry. Boschma, Eriksson and Lindgren (2009) found that the hiring of new employees with skills the plant already had inhouse had a negative impact on plant performance. This may be a result of the fact that the new skills could be quickly integrated into the routines of the plants, but they did not enhance overall productivity, and might even have caused conflict and rivalry with employees with identical skills. By contrast, the inflow of new employees with related (not similar) skills contributed positively to plant performance, possibly because it added new skills to the existing set of skills in the plant, and it formed no direct threat to employees with identical competences. The next step is to link this effect of post-entry labour mobility to the industrial life-cycle approach. This concerns questions like: what kind of labour, and what sets of skills (diversified, related, specialised) are needed for firms to survive over the different stages of the industry life-cycle?

An evolutionary approach has the potential to bring a new perspective on the study of regional clusters, because it meets to a considerable degree the critique advanced by Martin and Sunley (2003) and Brenner (2004), especially the tendency to treat clusters in too static a manner. In Chapter 8, Simona Iammarino and Phil McCann describe the evolutionary aspects of clusters (see also Iammarino, 2005; Menzel and Fornahl, 2009; Swann and Prevezer, 1996; Ter Wal and Boschma, 2009), and emphasise the importance of path-dependent processes that shape cluster development, and which make clusters look very different over time (see also Iammarino and McCann, 2006). In Chapter 10, Udo Staber provides a thorough critique on the cluster literature from an evolutionary angle. Staber argues that the cluster literature has underestimated the origins and consequences of (new) variation in clusters. Consequently, clusters have been misconceived as coherent entities while, in practice, they consist of different levels (e.g. individuals, ideas, routines, organisations, networks) on which selection operates. According to Staber, evolution proceeds at a faster pace at lower levels of action and, thus, the fitness of ideas increases more rapidly than the fitness of a cluster as a whole. Thus, while at the level of individuals, there may be much turbulence (such as exits and new network relationships), this will not necessarily affect the survival of the cluster as a whole. Research should focus more on identifying the dynamics of selection mechanisms (such as competition and imitation) at various levels to explain stability and change in clusters. This would meet worries expressed in the literature that clusters are often treated as static, instead of dynamic entities (Carlsson, 2003; Maggioni, 2002; Martin and Sunley, 2003). And cluster research has to recognise that clusters can develop dysfunctional features that often give rise to diseconomies and negative economic effects, a point also made by Martin and Sunley (2003). Staber proposes an alternative evolutionary view on regional clusters. His basic line of argument is that, while evolution may unfold very differently across clusters, the general Darwinian processes (variation, selection, retention) operate in all units and at all levels of actions.

Another critique on the cluster literature is that it almost ignores the existence and effects of collaborative relationships among constituent firms. In Chapter 11, Phil Cooke and Carla de Laurentis begin by explaining how an evolutionary view fundamentally differs from a neoclassical approach. They then present the results of an empirical research project on collaborators and non-collaborators in cluster and non-cluster settings in the UK ICT industry. The data show that collaborators in clusters perform better than non-collaborators in clusters, suggesting a premium effect of local networking. There is also some evidence that collaborators in clusters perform better than collaborators in non-cluster settings. However, non-collaborators in clusters perform worse than non-collaborators in non-clusters, suggesting a diseconomies of scale effect in clusters for non-collaborators. The study by Cooke and de Laurentis shows how the analysis of networks of collaboration is important for understanding the dynamic performance of clusters, and links to the topic that is the focus of the next part of the book.

4. Network evolution and geography

Part 3 of this book concentrates explicitly on the importance of networks in an evolutionary approach to economic geography. There is increasing awareness that networks play a crucial role in understanding the spatial uneven distribution of economic activity. Some have advocated the need for a relational turn in economic geography (Bathelt and Glückler, 2003; Boggs and Rantisi, 2003; for a critique, see Sunley, 2008). As explained in Boschma and Frenken (2006), networks can also be incorporated and analysed in an evolutionary framework. An evolutionary approach to networks can be applied to various types of spatial network, like infrastructure networks and urban networks (Taylor et al., 2007; Taylor and Aranya, 2008; Wall, 2009). In these latter cases, places are depicted as nodes in networks. In a seminal paper, Barabasi and Albert (1999) proposed that networks evolve as the result of an entry process of new nodes that connect with a certain probability to existing nodes, depending on the connectivity of the latter. This type of model can explain the emergence of 'hubs-and-spokes' structures in space, as found in airline networks, for example. Geographers are interested how location-specific characteristics and the geographical distance between new and existing nodes influence the formation of the network.

In this handbook, we limit our attention to inter-firm knowledge networks (Hagedoorn, 2002; Ozman, 2009; Powell et al., 1996). Only recently have geographers turned to the empirical study of the spatial dimensions of networks in innovation processes. These studies have spiralled out of the literature on national and regional innovation systems developed in the 1990s, which had strong evolutionary roots from the very start (Asheim and Gertler, 2005; Breschi and Malerba, 1997; Cooke, 1992; Edquist, 1997; Freeman, 1987; Nelson, 1993). The objective of the innovation system literature was to uncover the institutional setting in a territory that affects the interaction patterns between a range of organisations involved in the innovation process. This led to the insight that countries and regions have different innovation systems, the nature of which can only be understood by looking at their history, that is, how these systems were shaped and transformed over time. Sectoral studies of innovation systems (e.g. Malerba, 2002) have typically adopted such a dynamic perspective, setting out how institutions co-evolve with the emergence of a new sector (see, for example, Consoli, 2005; Murmann, 2003). Another promising line of research has focused on whether sectoral shifts in innovation lead to institutional changes at the national level, because of evolutionary forces like selection, retention and imitation of sector-specific institutional models (see Hollingsworth, 2000; Strambach, this volume Chapter 19), a topic to which we return in Part 4 of the book.

An evolutionary approach to spatial networks could contribute to the field of economic geography in at least three ways. First, the study of networks promises to provide additional insights in the workings of clusters (Giuliani, 2007; Uzzi, 1996). In an influential paper, Giuliani and Bell (2005) have applied social network ideas to demonstrate that knowledge networks in a cluster are not pervasive, as is often assumed by the cluster literature, but selective. The micro-evolutionary view developed by these authors links the network positions of firms in a cluster to their absorptive capacity. Second, we still have little understanding of what are the main drivers of network formation. Boschma and Frenken (Chapter 5) and Sorenson et al. (Chapter 15) claim that a proximity framework is useful in this regard. Such a framework enables us to isolate the effect of geographical proximity alongside other forms of proximity on network formation, because geographical proximity is just one potential driver, and not necessarily the most important one. This line of reasoning follows Boschma (2005), who claimed that geographical proximity is neither a sufficient nor a necessary condition for firms to engage in network relations and inter-firm learning. Third, the study of network dynamics and the role of geography is still in a premature stage, not least because of the limited availability of time series data on networks, and the embryonic methodological state of dynamic network analysis. Below, we elaborate on each of these three potential contributions more in detail.

Recent network studies have contributed to a better understanding of clusters (see, for example, Boschma and Ter Wal, 2007; Cantner and Graf, 2006; Guiliani, 2007; Giuliani and Bell, 2005; Morrison, 2008; Staber, 2001; Visser, 2009). Using a micro-perspective on knowledge networks, these studies have questioned the conventional view that the economic well-being of cluster firms ultimately depends on extra-firm sources of knowledge, rather than intra-firm routines. These social network studies have also challenged the view that knowledge is 'in the air' in a cluster, from which all cluster firms can equally benefit. Without exception, these studies show that only a limited number of firms in a cluster are well connected to the local knowledge network, while many cluster firms are poorly connected, or not connected at all. As Elisa Giuliani demonstrates in her study of three wine clusters in Chapter 12, this is not a trivial issue: a knowledge linkage between two cluster firms increases the likelihood that both firms perform well. So, it is not the cluster per se that matters for firm performance, but being connected to the local knowledge network. In addition, these network studies have shown that reliance on external knowledge relationships does not necessarily mean these are confined to the cluster. What these studies demonstrate is that the best performing firms tend to show a high connectivity to firms outside the cluster. A next question then is whether these leading firms might still act as gatekeepers for the cluster. Network studies show that this depends on their connectivity to other local firms, among other things (Morrison, 2008; Cantner and Graf, this volume Chapter 17).

Looking at the structure and nature of intra-cluster networks as a whole, in Chapter 12 Giuliani criticises the cluster literature for claiming that intra-cluster networks per se enhance economic development in the cluster as a whole. Her study on three wine clusters clearly shows that it depends on the structural properties of intra-cluster networks as to whether these will generate beneficial effects throughout the cluster. More specifically, her analyses show that this depends on how selective knowledge networks in clusters are, that is, how unevenly distributed these networks are among cluster firms. Giuliani also found that business networks (as opposed to knowledge networks) are more pervasive in clusters, but these do not spread positive spillovers throughout the cluster.

Concerning the firm level, network studies have recently taken up the question of what explains the network position of firms in clusters. In their seminal paper, Giuliani and Bell (2005) pointed to the importance of firm-specific competences, like the absorptive capacity of firms. In Chapter 13, Stefano Denicolai, Antonella Zucchella and Gabriele Cioccarelli draw attention to the formation of different forms of reputation and trust between local firms, and how these affect the formation and dissolution of network linkages at the cluster level. Instead of referring to firm-specific competences, Denicolai et al. describe how firms with different forms of reputation may occupy different network positions in clusters. They illustrate this with a number of network studies they conducted in three Italian clusters.

Instead of focusing on clusters, network studies have also investigated the drivers of network formation at the level of firms. In Chapter 14, Johannes Glückler examines what drives the formation of sales partnerships in the stock photography sector in Germany. Based on a network survey of firms, Glückler explains the likelihood of establishing a sales partnership between two agents. Borrowing network measures from social network analysis, his findings suggest that the likelihood of forming a sales partnership in this sector is primarily shaped by multi-connectivity, rather than geographical proximity and homophily, for instance. In Chapter 15, Olav Sorenson, Jan Rivkin and Lee Fleming analyse US patent data and citation rates across inventors to determine which forms of proximity are important for knowledge flows, and relate these to the nature of knowledge (whether complex or not). Synthesising a social network view with a perspective of knowledge transfer as a search process, they claim that the advantages of being proximate to some knowledge source depend crucially on the nature of the knowledge at hand. Their findings show that simple knowledge flows equally to actors near and far, while complex knowledge is unlikely to diffuse, no matter how proximate actors are. With knowledge of moderate complexity, however, the outcomes show that closer actors are in a better position to benefit from knowledge diffusion, in contrast to more distant recipients. Sorenson et al. conclude that an interesting line of research would be to investigate how the tendency of industries to agglomerate might depend on the complexity of knowledge.

Regional networks may be formed through the movement of labour. Labour mobility has the potential to provide effective channels of knowledge diffusion across countries and regions (Saxenian, 2006). Agrawal et al. (2006) found that knowledge is transferred between firms across large geographical distances when their respective employees are socially linked because of a shared past in the same school or same company. In other words, knowledge networks are formed through social proximity between agents, irrespective of their geographical distance, although social networks are often geographically bounded (Breschi and Lissoni, 2003). In Chapter 16, Stefano Breschi, Camilla Lenzi, Francesco Lissoni and Andrea Vezzulli explore a large set of patent applications by US inventors registered at the European Patent Office in three high-technology fields in the period 1991 to 1999. They demonstrate that inventors who patent across firms do not diffuse their knowledge that much across space, first, because inter-regional mobility of inventors is rather limited, and second, because inventors create social networks at the regional level, not across regions. The few inventors who do move between regions, however, tend to maintain their ties with former co-inventors, providing a channel of knowledge diffusion to their prior location. In the latter case, knowledge diffuses across space through the professional networks of inventors.

Another issue is the effect of (regional) networks on economic performance. Studies often report a positive relationship (Ozman, 2009). However, this is not necessarily the case. This may depend on, among other things, the degree of proximity between the networks partners. While a high degree of any form of proximity might be considered a prerequisite to make agents interact, proximity between agents does not necessarily increase their innovative performance, and may possibly even harm it (Boschma, 2005; Broekel and Meder, 2008; Cantner and Meder, 2007; Grabher and Stark, 1997; Nooteboom et al., 2007). This is in line with what Gilsing et al. (2007) found when they assessed the impact of technological distance in high-tech alliance networks on the innovative performance of the partner firms. As expected, they found an inverse U-shaped function between technological distance and exploration (i.e. real breakthroughs), meaning that neither a very high nor a very low degree of technological proximity between partners resulted in exploration.

As Boschma and Frenken set out in Chapter 5, the study of dynamic spatial networks has thus far been limited, because of the limited availability of longitudinal data on networks and the poor methodological state of dynamic network analysis (see Ter Wal, 2009). Having said that, in the context of evolutionary economic geography, the main challenge is the study of the dynamics of network formation: how do networks of firms arise and develop in time and space, and what forms of proximity are important at what stage of the evolution of the network? In this context, the focus of attention is on the dynamics in the number of nodes and relations, and how the different forms of proximity impact on these network dynamics. What is interesting from an evolutionary perspective is to examine whether the different proximities induce path dependence in network evolution, and whether they cause retention in the local network (Glückler, 2007). Moreover, a dynamic network approach should account for that fact that the evolution of a network structure may, in turn, affect the degree of proximity in its various dimensions, like in the social and cognitive dimension (Menzel, 2008). However, this requires further refinements at the conceptual and methodological level before this can be applied empirically (see for example Glückler, Chapter 14).

In Chapter 17, Uwe Cantner and Holger Graf apply a dynamic perspective on regional innovation systems. Their contribution provides a study on the evolution of the Jena innovation network based on patent data. Behind the increase of the network size and the degree of connectedness, considerable dynamics are observed in terms of actors entering and exiting the network. They demonstrate that permanent innovators and new entrants show a tendency to concentrate more on the technological core competences of the network, while the network as a whole shows an increasing inward orientation to the local level. This chapter gives evidence of the enormous potential of analysing network dynamics at the regional and local scales, a research field that is still underdeveloped.

Taking a dynamic perspective on network evolution, one can also theorise from the industry lifecycle concept about the role of inter-firm networking at different stages of the lifecycle (see Boschma and Frenken, Chapter 5; Ter Wal, 2009). This is not to say that each industry is destined to follow such a life-cycle, and that each stage follows the other in a deterministic manner. Already in the 1980s, some economic geographers challenged the application of the industry lifecycle approach in such a stylised manner (see for example Taylor, 1986). Moreover, this would go against the nature of economic evolution as an open-ended process that is conditioned but not determined by its spatial context (Martin and Sunley, 2006). However, the main purpose of developing an endogenous model of network evolution along industry life-cycle lines is to derive some hypotheses that can be tested case by case, in order to determine the empirical veracity and applicability of the ideal type, especially in different spatial contexts. Ter Wal and Boschma (2009) have applied an industry life-cycle framework to the study of network dynamics. Following the life-cycle scheme of Abernathy and Utterback (1978), they expect that firms engage in network activity in the first phase to explore the possibilities of a new technology. As technologies are still competing and uncertainty is high, the network is unstable, with firms often changing partners. In the second phase, firms developing dominant designs become the hubs while lagging firms try to maintain their position by networking with a hub (see Orsenigo et al., 1998). In the mature phase, the degree of embeddedness in networks is high as peripheral firms have exited. This can lead to over-embeddedness and 'lock-in' (Grabher, 1993). What is more, the repeated interactions in the past might have increased the cognitive proximity between firms, thereby reducing the potential for innovation by recombination. In certain circumstances, this can lead to an endogenous decline of the cluster (see also Menzel and Fornahl, 2009; Staber, 2007). For a discussion of how positive lock-in can turn into negative lock-in, see also Martin and Sunley (2006) and Hassink (Chapter 21). Institutions also play a prominent part in this respect, a topic to which we turn now.

5. Institutions, co-evolution and economic geography

Since its early development, institutions have been part and parcel of evolutionary economics. Veblen, one of its founding fathers, emphasised the importance of habits, conventions and norms in economics (1898). He laid the foundations of what is now called 'old' institutionalism, and which has affinities with evolutionary economics as this has developed since the 1980s. In the late 1970s, Nelson and Winter (1977) developed their concept of 'natural trajectories', which were described as heuristics that guide the innovation process (see also Rosenberg, 1976). These were often driven by the logic of the mechanisation and standardisation of production, with the purpose of constraining wages by codifying the tacit knowledge of employees. In this way, Nelson and Winter gave their evolutionary theory a kind of Marxist flavour by referring to the overall importance of capital–labour conflicts. In the 1980s, the so-called regulation theory (Boyer, 1988) held a prominent place in the seminal contribution of Dosi et al. (1988) on the foundations of evolutionary economics. Moreover, eminent evolutionary scholars like Freeman and Perez (1988) developed the concept of structural crisis of adjustment, in which they claimed that institutions need to be transformed to enable new industries to develop fully and for old industries to be revived. And in the late 1980s, the innovation system literature was initiated and developed by prominent evolutionary economists like Freeman (1987), Lundvall (1992), Nelson (1993, 2002) and Malerba (2002, 2004) who likewise emphasised the importance of national and sectoral institutions for the innovation process.

Economic geographers have been keen to apply these ideas to their own discipline. In the 1980s and 1990s, some adopted a regional approach to regulation theory, but this was never fully developed (Martin, 2000). More successful has been the adoption of the concept of innovation systems, which was explored in the early 1990s (see, for example, Asheim and Gertler, 2005; Cooke, 1992, 2001; Cooke et al., 1998), and to which economic geographers continue to contribute to this very day. Interestingly, evolutionary economists introduced the concept of innovation systems by linking it to a particular geographical scale right from the start, that is, the national dimension (Freeman, 1987; Nelson, 1993). What economic geographers have added in the meantime is the view that innovation processes are firmly rooted in region-specific institutions, most them of an informal nature (e.g. local culture) that are difficult to copy or imitate by actors in other regions. Because of these intangible assets, regions are considered important drivers of innovation, despite tendencies of globalisation (Belussi and Sammarra, 2005; Storper, 1992).

In their attempt to delineate an evolutionary approach in economic geography, Boschma and Frenken (2006) made the observation that institutions have not always been treated by economic geographers in a truly evolutionary manner. Broadly speaking, they came to four lines of criticism: (1) institutions are often presented as pre-given and fixed, as if these come from nowhere (a-historical) and do not change over time; (2) institutions are often depicted as factors that determine, instead of condition, the economic behaviour of agents and the performance of regions; (3) institutional approaches in economic geography tend to employ case-study approaches while ignoring, if not rejecting altogether, the use of quantitative methodologies. Although case studies have generated many valuable insights into processes of regional innovation, many institutional approaches have been reluctant to test any hypotheses that might be derived from these; (4) institutional approaches in economic geography tend to associate institutions with territories (at whatever spatial level) and, hence, spatial differences in economic activities are attributed to institutional differences among territories. This stands in contrast to those evolutionary approaches to economic geography that reason from organisational routines, and that view the behaviour of firms as mainly stemming from their routines, rather than from territorial institutions.

When taking such a firm-based, micro-perspective, one avoids running the risk of over-emphasising the role of territorial institutions and violating a crucial ingredient of an evolutionary approach, that is, the heterogeneity of firms. The assumption that all agents act or perform the same when subject to the same regional institutions also contradicts empirical findings that suggest the opposite. As noticed above, Giuliani (2007), among others, has demonstrated that agents in clusters differ widely in terms of economic power, absorptive capacity and network position, despite the fact that clusters are associated with a particular set of institutions. This variety can only be understood from the fact that firms develop routines in a path-dependent and idiosyncratic manner, and territorial institutions are often so general such that specific effects at the firm level can still vary greatly (Boschma and Frenken, 2009a). The very fact that a variety of routines is found in a territory shows that territory-specific institutions do not determine the kinds of routine that develop and survive in a territory. As Gertler (2009) puts it, there is 'a danger of "reading off" individual behaviour from territorial institutions'. Instead, there is a need to account for the role of contingency in regional development (see e.g. Bathelt and Glückler, 2003). This implies taking the individual and firm level more seriously than institutional studies have tended to in the past.

This is not to deny that territorial institutions may explain part of the inter-regional variety of routines, however. For example, production techniques of plants in a number of manufacturing industries have been found to be more similar within than across US regions, and that these regional differences are quite persistent over time (Essletzbichler and Rigby, 2005; Rigby and Essletzbichler, 1997). This may be attributable to region-specific institutions, but it may also be the result of routine replication among local firms through spinoff dynamics or labour mobility effects. Therefore, we have to be cautious about taking the impact of institutions for granted, and assess their relative importance on a case by case basis (Boschma and Frenken, 2009a).

To incorporate institutions more fully into an evolutionary economic geography framework, MacKinnon et al. (2009) claim there is a need to account for power and labour–capital conflicts. Boschma and Frenken (2009a) have argued that this can be accomplished when accounting for the political dimension of routines, as advocated by Nelson and Winter (1982). Besides the cognitive dimension of routines, which has drawn most attention among evolutionary scholars, Nelson and Winter (1982) have defined routines as a mechanism of internal control, more specifically to control potential labour–capital conflicts within the firm (see also Nelson and Winter, 1977; Rosenberg, 1969). Integrating the political aspect of routines in an evolutionary approach to economic geography would start from the study of how firms regulate or resolve conflicts of interests between capital and labour differently using different routines. A geographical perspective should aim then to explain the diffusion of such routines among firms within and across regions, and to determine under what conditions such a diffusion process leads to an institutionalisation of particular routines at specific territorial levels (Boschma and Frenken, 2009a).

Quite recently, there has been increasing attention directed to how institutions can be incorporated in the explanatory framework of evolutionary economic geography (see, for example Boschma and Frenken, 2009a; Essletzbichler, 2009; Grabher, 2009; Hodgson, 2009; MacKinnon et al., 2009). Part 4 of this volume brings together contributions of leading economic geographers that take up this major challenge. What they tend to criticise is that institutions are often treated as static entities that are left unexplained. In Chapter 18, Anders Malmberg and Peter Maskell combine a micro-perspective with a dynamic macro-view on institutions in economic geography. According to Malmberg and Maskell, there is a strong need to understand better how institutional dynamics form and shape particular paths at the aggregate level of regions and countries. Linking micro-behaviour to macro-level processes, they explore how these evolutionary insights might be applied to the analysis of the birth, growth and decline of clusters, a topic that has been largely ignored by the cluster literature. In this respect, they develop a stylised version of an evolutionary approach to the cluster life-cycle.

In Chapter 19, Simone Strambach focuses on how new technological development paths emerge within established institutional settings and create institutional change. She goes beyond the common use of the notion of path dependency that merely emphasises the stabilising and action-guiding functions of institutions. Interestingly, Strambach takes the view that institutional systems are not necessarily coherent themselves, but subject to institutional *plasticity*, meaning that a range of options for new development paths are open within the overall dominant institutional system. Creative agents can deviate from the established path in a deliberate and purposeful manner, creating new institutions but not necessarily breaking with the existing institutional system (see also Martin, 2010b). Strambach takes the rise of the German customised business software industry as an example to demonstrate that a new path can be created within an unfavourable and incompatible institutional setting. Despite the hostile German national innovation system, this sector succeeded to become a highly competitive sector on the world market, creating new supportive institutions and adapting established ones. In other words, Simone Strambach proposes an evolutionary approach that aims to endogenise the role of institutions, and makes institutions a more integral part of the explanation of the evolution of the economic landscape.

This echoes the view of Nelson (1995) who proposed that institutions should be thought of as co-evolving with technology and markets. There are some fine studies that focus on the interplay between industrial and institutional change, by describing how institutions co-evolve alongside new industries (Freeman and Perez, 1988; Murmann, 2003). This literature emphasises that supportive institutions are often an outgrowth of emerging industries, but once they become established, institutions may obstruct their further development because of inertia and hysteresis (Martin and Sunley, 2006; Setterfield, 1997).[5] Institutional change is therefore required not only to enable the emergence of new industries, but also to revive mature industries. We still have little understanding of the conditions under which regions or countries are more likely to adapt their institutions to seize opportunities provided by new sectors, and under what conditions institutional adaptation fails to take place (Maskell and Malmberg, 2007).

In Chapter 20, Eike Schamp takes up this issue but in a much broader sense. He explores how the notion of co-evolution may be fruitfully applied in evolutionary economic geography. According to Schamp, the notion of co-evolution has often been used in economic geography in a rather narrative and loose sense. Following Malerba's work, he argues there is a strong need to identify what is co-evolving, and to be more specific about the reciprocal causalities between the co-evolving entities that are analysed. In this context, Schamp discusses several empirical topics that focus on co-evolutionary processes: the simultaneous growth of connected sectors, the co-evolution of sectors and institutions, and the creation of a region as a supporting environment.

In this context, Robert Hassink explores in Chapter 21 the relationship between lock-in and regional development. While studies in economic geography tend to emphasise the positive effects of geographical clustering on growth, Hassink elaborates on the idea that clusters may also end up in a state of negative lock-in (see also Hassink, 2005; Hassink and Shin, 2005; Schamp, 2005). In his contribution, Hassink makes the case that the emergence and persistence of these negative effects can be well explained by an evolutionary economic geography approach. Empirical cases of divergent experiences in old industrial regions in Germany and South Korea are discussed to show: (1) why in some old industrial regions strong lock-ins are found, and in other industrial areas weak lock-ins; and (2) how regional lock-ins may be caused by (institutional) structures at different spatial levels, including the national and supra-national level.

While many institutions are heavily influenced by the state, what is still missing though is a systematic view on the state that is well grounded in evolutionary thinking (see, for example, Metcalfe, 1994; Pelikan and Wegner, 2003). To an increasing extent, there are contributions that specify how evolutionary economic geography may inform regional innovation policy (see, among others, Asheim et al., 2006; Boschma, 2009; Lambooy and Boschma, 2001; Nooteboom and Stam, 2008; Raspe and Van Oort, 2006; Shapiro and Fuchs, 2005; Tödtling and Trippl, 2005; Van Geenhuizen and Nijkamp, 2006). What these approaches tend to argue is that region-specific contexts provide opportunities but also set limits to what can be achieved by public policy. Consequently, policy action should avoid 'one-size-fits-all' and 'picking-the-winner' policies. Instead of copying best practices or selecting winners, policy should take the history of a region as a starting point, and identify regional potentials and bottlenecks accordingly. To avoid the problem of regional lock-in, Nooteboom and Stam (2008), among others, have argued that public policy should stimulate the entry of newcomers, encourage new policy experiments, and enhance the establishment of extra-regional linkages.

Another pending issue is how to incorporate institutions in a more quantitative evolutionary framework. Institutions have been mainly (though not entirely) examined in economic geography by using qualitative, descriptive case studies. Boschma and Frenken (2009a) have expressed the need for studies on institutions and regional development that strive for generalisations that go beyond the unique. This would make empirical studies in evolutionary economic geography 'more comparable, transparent and cumulative', without denying the importance of qualitative studies (see Markusen, 1999). Having said that, we are in need of more sophisticated methodologies that can cope with the explicit dynamic nature of evolutionary processes in economic geography (Boschma and Martin, 2007). Many approaches are being pursued to this end, ranging from case-study research and network analysis, to duration models, simulation methods and the use of spatial econometrics, depending on the research questions and the data available (Frenken, 2007). This methodological openness or pluralism may be considered a strength of evolutionary economic geography, as compared to neoclassical- and institutional-based approaches (Boschma and Frenken, 2006).

6. Structural change, agglomeration externalities and regional branching

Schumpeter once stated that economic growth is not just about quantitative change, but also about qualitative change. Long-term economic growth depends on the ability of countries to create new variety through entrepreneurship and innovation, in order to offset decline in other parts of the economy. Schumpeter conceived this process of 'creative destruction' as the driving force of economic development. Since the reappraisal of Schumpeter's work in the late 1970s, economic geographers have applied these Schumpeterian ideas on structural change and industrial dynamics to regional development in a variety of contexts (e.g. Hall and Preston, 1988; Lambooy, 1984; Marshall, 1987; Norton, 1979; Norton and Rees, 1979; van Duijn and Lambooy, 1982). In the 1980s, there was an almost general consensus among economic geographers that 'new industries' do not emerge in 'old regions'. Empirical studies showed that new industries flourished in new growth regions like the Sunbelt states in the US, the south east in the UK and Bavaria in Germany, while old and declining sectors were mainly located in what were once the leading regions in the US (like the Rustbelt) and Europe (like the North in the UK, and the Ruhr area in Germany). Conventional approaches took a deterministic view on this, claiming that new sectors had different locational demands, like quality of life and low (labour) costs, as compared to old sectors. Other approaches took a more evolutionary perspective, emphasising that it is unpredictable where new growth industries will emerge and change the economic landscape, because of chance events in combination with increasing returns (Boschma and Lambooy, 1999; Storper and Walker, 1989).

The spatial evolution of the economic system at the macro level may be addressed in a framework of structural change, in which catching-up and falling-behind of regions is analysed not only in terms of the rise and fall of sectors but also in terms of the rise and decline of (infrastructure) networks. In this respect, the economic development of cities and regions can be analysed as an aggregate of sectoral change, and from their (changing) position in global networks of trade and knowledge. With respect to sectors, cities and regions that are capable of generating sectors at the start of a product life-cycle will experience growth, while cities and regions that are locked into mature stages of life-cycles will experience relative decline. There is no automatic economic or political mechanism that assures cities or regions will successfully renew themselves in this respect (Boschma and Frenken, 2006). With respect to networks, the growth of cities and regions also depends on a city's or region's inclusion in global networks of trade and commerce

(Castells, 1996; Hohenberg and Lees, 1995). A central network position can be achieved by attracting corporate headquarters, developing specialised business services, and functioning as major transportation hubs. On the one hand, one might expect cities and regions in one historical era (e.g. based on railways) to be less successful in the next era (e.g. based on airlines), because of institutional rigidities and sunk costs associated with previous infrastructures. On the other hand, some major cities (like London and New York) seem to be capable of maintaining their leading positions in world-wide operating networks (Wall, 2009).[6]

The fifth and final part of this volume focuses on the relationship between structural change and the evolution of the spatial system at the macro level. Jan Lambooy, being one of the founding fathers of evolutionary economic geography, takes up this issue in Chapter 22. He explores how an evolutionary approach in economic geography may deal with the inter-relationship between structural change and the evolution of the economic landscape. While the relation between technology and economic development has drawn a lot of attention, Lambooy argues that the impact of this relation on spatial structures, in particular urbanisation, has remained relatively unexplored. He claims that spatial structures tend to reflect technological and economic development in various ways, but often with a time lag, because of physical and institutional constraints that are engraved in space. More particularly, Lambooy discusses how general purpose technologies like ICT have impacted on spatial patterns, such as the process of urbanisation and the spatial evolution of industries and networks.

Evolutionary economic geography deals with the uneven distribution of economic activity across space, and how that evolves over time. In Chapter 23, James Simmie takes up how new technological regimes impact on the evolution of the economic landscape. He examines the example of the rise of the service-based economy, and he addresses the need for an evolutionary perspective to investigate its spatial implications. Drawing on recent experiences in the evolution of the English urban system, Simmie investigates two recent phenomena in the rise of the information society, namely the rise of knowledge-intensive business services and the importance of network ties between service sectors in the Greater South East region.

An evolutionary approach centres on historical processes that produce the uneven economic landscape. In this respect, spatial patterns emerge from economic growth processes that occurred in the past (Simmie and Carpenter, 2007). At the same time, spatial distributions affect subsequent patterns of growth because of the uneven spatial distribution of resources built up in the past giving rise to (positive and negative) externalities. Stochastic models of urban growth using time series on city size investigate sustained urban growth and decline, thus going beyond the logic of Gibrat's Law stating that urban growth rates are stochastic and independent of city size (Pumain and Moriconi-Ebrard, 1997). This approach falls under evolutionary economic geography, since these models account for path dependence in which each event changes the probability of a next event to occur (Arthur, 1989; David, 1985). Such an evolutionary perspective differs from the core model in new economic geography (NEG) (Krugman, 1991) where changes in spatial distributions are explained from parametric changes, as in transport cost. The concept of path dependence in that model is different in that it refers to multiple equilibria that are sensitive to initial conditions only (Boschma and Frenken, 2006; Martin, 1999, 2010a, 2010b). What unites evolutionary models is that the growth dynamics are path-dependent, and that this path dependence does not simply arise from the assumption of increasing returns as is the case in NEG models. That is not to say that increasing returns do not play a role. Rather, if included one should specify both positive and negative externalities.

In Chapter 24, Giulio Bottazzi and Pietro Dindo provide a fine example of how modelling may contribute to the further advancement of evolutionary economic geography. These authors explain how different their evolutionary model of firm location is from the neoclassical Krugman model that laid the foundations of the new economic geography. They present the outlines of an evolutionary entry–exit model of firms' location that describes how the economic landscape evolves over time. As a starting point, they present a static framework in which a spillover drive scenario leads to agglomeration, while a market drive scenario may generate either agglomeration or even spatial distribution. The model further specifies how other variables (like transport costs) may enhance technological spillovers or market forces and thus agglomerative forces, or not. This static framework is complemented by an evolutionary entry–exit model, in which heterogeneous firms may change their locational preferences, because of previous decisions of other firms. In this dynamic setting, agglomeration is a less likely outcome, and when it occurs, the agglomeration may not always be stable.

Another promising line of research in evolutionary economic geography is to determine what kind of agglomeration externalities are needed to promote urban and regional growth (Feldman and Audretsch, 1999; Glaeser et al., 1992; Henderson et al., 1995; Jacobs, 1969). Frenken et al. (2007) have gone beyond the dichotomy of Marshall-Arrow-Romer (MAR)-type versus Jacobs-type externalities by introducing the notion of 'related variety' type externalities. This means that regions that are endowed with technologically related sectors might have higher growth rates, because this might affect positively the nature and scope of regional knowledge spillovers. That is, the extent to which the variety of technologies present in a region are related is expected to affect the scope for knowledge spillovers, as firms in different but related activities can profit more from mutual spillovers than can firms in unrelated activities (Boschma and Frenken, 2009b). In other words, related variety performs two tasks at the same time. Some degree of cognitive proximity (i.e. relatedness between sectors) ensures that effective communication and interactive learning between sectors take place. But also some degree of cognitive distance (that is, variety between sectors) is needed, to avoid cognitive lock-in and stimulate novelty (Nooteboom, 2000). Frenken et al. (2007) could demonstrate empirically for the Netherlands that related variety does indeed have a positive impact on regional growth. This result has been replicated in studies on other countries (Bishop and Gripaios, 2009; Boschma and Iammarino, 2009; Essletzbichler, 2007).

The next step to take in these regional growth models is to account for the fact that new and related variety may also be brought into the region through inter-sectoral linkages with other regions. Boschma and Iammarino (2009) have made a first attempt to estimate the effects of inter-sectoral learning across regions on regional growth in Italy by means of trade data. Their analysis suggests that the inflow of a variety of knowledge per se did not affect economic growth of regions in the period 1995–2003. The same was true when the extra-regional knowledge was similar to the knowledge base of the region. However, the more related the knowledge base of the region and its import profile was, the more it contributed to regional employment growth. This might indicate that a region benefits especially from extra-regional knowledge when it originates from sectors that are related or close to, but not quite the same as the sectors present in the region. This type of analysis goes beyond the regional level, and sheds more light on the importance of inflows of extra-regional knowledge (is it just a matter of being globally connected, or is there more to say?). However, more refinement is needed to assess more systematically its importance for urban and regional growth.

This also concerns how to measure related variety. Recently, scholars have come up with more sophisticated indicators of relatedness on the basis of combinations of human skills or products that occur frequently in plants or firms (Breschi et al., 2003; Hausmann and Klinger, 2007; Neffke and Svensson Henning, 2009; see for a discussion Neffke, 2009). Such studies provide a picture of which industries are related to one another, in order to capture better the knowledge spillover effects of related variety. Another advantage of these new indicators of relatedness is that they do not rely on predefined and static SIC codes. Since relatedness between industries may change in the long run because of technological developments, there is a need for a flexible indicator that accounts for shifts in technological relatedness and related variety over time.

The agglomeration economies literature (Henderson et al., 1995) has claimed that new (high-tech) industries need Jacobs-type externalities (and thus inter-industry knowledge spillovers) to develop, while more mature industries benefit more from MAR-type externalities (i.e. intra-industry spillovers) in more specialized cities. From an evolutionary point of view, it is more interesting to investigate whether new industries need the local presence of related industries. Following the idea behind related variety, we would expect that a local diversity of sectors per se is less likely to lead to successful new combinations, because sectors will learn more from each other when they are technologically related. A research question that follows from this is whether related variety itself can be explained as an outcome of a historical process of regional development. As Boschma and Frenken (2009b) have explained, different time scales are involved. In the short run, related variety is a very stable property as the sectoral composition of a regional economy changes only slowly over time. Yet, over longer time scales, related variety itself is subject to change and becomes a dependent variable. One can ask the question of to what extent the technological relatedness between sectors in a regional economy as a whole can help us to understand the opportunities of each region to diversify into new and related industries. To the extent that new industries emerge from existing and related industries, the sectoral composition of a regional economy at one moment in time provides but also constrains diversification opportunities of regions in the near future.

Thus, from an evolutionary economic geography perspective, one would expect that a set of related industries in a region is rather persistent over time because regions are more likely to expand and diversify into sectors that are closely related to their existing activities (Hidalgo et al., 2007; Neffke and Svensson Henning, 2009). This means that when firms diversify (but not many do so because of the risks involved), they will show a higher propensity to diversify into technologically related instead of unrelated industries. Recent studies have indeed confirmed that regional branching occurs through related industries. Hausmann and Klinger (2007) found empirical evidence that there is a strong tendency of the export mix of countries to move from current products towards related products, rather than goods that are less related. In other words, a country's current position in the product space determines its opportunities for future diversification.

Thus the process of structural change is very much conditioned by existing related activities in a territory, providing support for spatial path dependence. Neffke and Svensson Henning (2009) found evidence that unrelated sectors are more likely to exit the region than related sectors, while sectors that are related to other sectors in the regional portfolio are more likely to enter the region, as compared to unrelated sectors. So, regions might change their industrial profile over time, but they tend to do so in a slow manner, and when they diversify, this is rooted in their existing industrial profile (Neffke, 2009). However, this is not to say that every country or region has the same probability to diversify successfully into related activities. Hausmann and Klinger (2007) found in their study that rich countries that are specialised in the more dense parts of the product space, have far greater opportunities to sustain economic growth, as compared to poorer countries that are positioned in the less dense parts. In conclusion, the historical trajectories of regions shape the rise and fall of sectors, but are also in turn shaped and transformed by this process of creative destruction.

More research is needed, but these outcomes suggest that the long-term development of regions depends on their ability to develop new sectors or new market niches that have their roots in the current regional knowledge base. It means that regional economies should branch out in new directions rather than start from scratch when they diversify. Frenken and Boschma (2007) and Boschma and Frenken (2009b) have suggested that branching occurs at the regional level because it becomes manifest through knowledge transfer mechanisms such as spinoff activity, firm diversification, labor mobility and networking, all of which tend to be geographically bounded. This opens up a whole new research agenda.

7. Postscript: looking ahead

Our aim in this introductory chapter has been to establish the case for, and to map out the scope of, an evolutionary perspective in economic geography. We would not be so bold as to claim that this handbook offers a definitive and fully articulated theoretical and methodological framework for an 'evolutionary economic geography'. After all, even evolutionary economics, which has been developing in its modern guise for some three decades or more, has yet to achieve a 'stable shared meaning' (Klaes, 2004). Nevertheless, it should be clear from our foregoing discussion and survey that the body of literature exploring, embodying and employing evolutionary metaphors, ideas and concepts in the analysis of the spatial economy has begun to grow apace, and that some of the main outlines of what is arguably a new paradigm in economic geography are already taking shape. The contributions to this book provide the first major collective statement on the nature and possible future direction of this new paradigm.

The construction of a coherent evolutionary economic geography is very much a 'work in progress'. Like all new paradigms in their early stages of development, the focus has been on what we might call 'abduction', that is the creative transfer of metaphors from one scientific discourse to another, in the case of evolutionary economic geography the key metaphors from evolutionary sciences – such as variety, selection, adaptation, and self-organisation. Used judicially, metaphors can have highly novel and creative effects: they can suggest entirely new ways of approaching the study of a phenomenon, and may point to all sorts of processes and effects not previously considered. The application and use of evolutionary metaphors in economic geography is having precisely these impacts.

Of course, as we have noted, the use of biological metaphors in economic geography raises it own set of issues, some of which are highlighted in the contributions to this handbook. Metaphors, by their nature are never complete, precise or literal mappings. If they were precise representations they would not be metaphors, and the juxtaposition of similar and different conceptual frameworks would be lost. The ontological challenge is to recast the evolutionary metaphors abducted from biology, the theory of complex adaptive systems and other such fields, and imbue them with geographical content and meaning, with the aim of eventually constructing a theory of spatial economic evolution.

This is not to suggest that an evolutionary perspective will or should replace other approaches to economic geography. We do not believe in theoretical or methodological monism. As a discipline, economic geography is necessarily and rightly eclectic and pluralistic in theoretical and empirical orientation. But we do believe that an evolutionary perspective has much to contribute to our discipline: it offers a distinctive way of looking at the economic landscape, it provides new insights, and it helps to stimulate an appreciation of historical change more widely within the subject. We hope this handbook will aid the advance of evolutionary economic geography on all three of these fronts.

Notes

1. Almost all of the contributions to the book started life as, or are based on, presentations made at this workshop. The support of the ESF is gratefully acknowledged.
2. A recent book by Jovanovic (2009) entitled *Evolutionary Economic Geography* is partial in its conception and coverage, and indeed seems to blend evolutionary economic geography with elements of the 'new economic geography', which to our mind is a rather forced – and incompatible – marriage of perspectives.
3. In fact a fourth field can be identified, namely that of panarchy (see Gunderson and Holling, 2002), in which the focus is on the evolutionary-type notions of adaptive cycles and resilience in ecological and social systems. These ideas have yet to be taken up by evolutionary economists and economic geographers, though they are beginning to attract interest (e.g. Hill et al., 2008).
4. This echoes recent calls from economic geographers to advocate approaches in which prime attention is given to the firm rather than the region (e.g. Boschma 2004; Taylor and Asheim 2001; Malmberg and Maskell, Chapter 18, this volume). After having reviewed a number of theories of the firm, including agency theory and transaction cost economics, Maskell (2001a) claims that evolutionary economics is especially useful to economic geographers because of its emphasis on learning and innovation, and because of the possibility to analyse territorial aggregates of firms in regional and national innovation systems (see also Maskell, 2001b).
5. Though it should be noted that Setterfield (1997) argued for a model of path-dependent economic development (and cumulative causation) that endogenised the evolution of institutions. In his model, institutions are regarded as fixed (exogenous) in the short run (such stability being necessary for economic growth and accumulation to proceed in confidence), but as changeable (and endogenous) in the medium to longer run, responding to the changing imperatives and needs of the economy.
6. Glaeser's (2005) study of how Boston has repeatedly 'reinvented' itself over the past three hundred and fifty years provides a striking example of how a city can overcome institutional and sunk costs, and renew its growth dynamic, in Boston's case by drawing on and replenishing its skilled labour force.

References

Abernathy, W.J. and J.M. Utterback (1978), 'Patterns of industrial innovation', *Technology Review*, **80**, 41–47.

Agrawal, A., I. Cockburn and J. McHale (2006), 'Gone but not forgotten: knowledge flows, labor mobility, and enduring social relationships', *Journal of Economic Geography*, **6**, 571–91.

Amin, A. and N. Thrift (2000), 'What kind of economic theory for what kind of economic geography?', *Antipode*, **32** (1), 4–9.

Arthur, W.B. (1989), 'Competing technologies, increasing returns, and lock-in by historical events', *The Economic Journal*, **99**, 116–31.

Arthur, W.B. (1994), *Increasing Returns and Path Dependence in the Economy*, Ann Arbor, MI: University of Michigan Press.
Arthur, W.B., S. Durlauf and D. Lane (eds) (1997), *The Economy as a Complex Evolving System, II*, Reading, MA: Perseus Books.
Asheim, B.T. and M.S. Gertler (2005), 'The geography of innovation: regional innovation systems', in J. Fagerberg, D.C. Mowery and R.R. Nelson (eds), *The Oxford Handbook of Innovation*, Oxford: Oxford University Press, pp. 291–317.
Asheim, B., R.A. Boschma, P. Cooke, A. Dahlstrand-Lindholm, P. Laredo and A. Piccauga (2006), *Constructing Regional Advantage. Principles, Perspectives, Policies*, DG Research, European Commission, 1–90.
Bagchi-Sen, S, and H. Lawton Smith (eds) (2006), *Economic Geography: Past, Present and Future*, Routledge Studies in Economic Geography, London: Routledge.
Barabasi, A.L. and R. Albert (1999), 'Emergence of scaling in random networks', *Science*, **286** (5439), 509–12.
Bathelt, H. and J.S. Boggs (2003), 'Towards a reconceptualization of regional development paths: is Leipzig's media cluster a continuation of or a rupture with the past?', *Economic Geography*, **79**, 265–93.
Bathelt, H. and J. Glückler (2003), 'Towards a relational economic geography, *Journal of Economic Geography*, **3**, (2), 117–44.
Beinhocker, E.D. (2006), *The Origin of Wealth: Evolution, Complexity and the Radical Remaking of Economics*, London: Random House.
Belussi, F. and A. Sammarra (eds) (2005), *Industrial Districts, Relocation, and the Governance of the Global Value Chain*, Padova: Università di Padova.
Bishop, P. and P. Gripaios (2009), 'Spatial externalities, relatedness and sector employment growth in Great Britain', *Regional Studies*, **4** (44), 443–54.
Boggs, J.S. and N.M. Rantisi (2003), 'The "relational turn" in economic geography', *Journal of Economic Geography*, **3**, 109–16.
Boschma, R.A. (2004), 'Competitiveness of regions from an evolutionary perspective', *Regional Studies*, **38** (9), 1001–14.
Boschma, R.A. (2005), 'Proximity and innovation. A critical assessment', *Regional Studies*, **39** (1), 61–74.
Boschma, R. (2007), 'Path creation, path dependence and regional development', in J. Simmie and J. Carpenter (eds), *Path Dependence and the Evolution of City Regional Development*, Chapter 3, Working Paper Series No. 197, Oxford: Oxford Brookes University, pp. 40–55.
Boschma, R.A. (2009), *Evolutionary Economic Geography and its Implications for Regional Innovation Policy*, Paris: OECD.
Boschma, R.A. and K. Frenken (2003), 'Evolutionary economics and industry location', *International Review for Regional Research*, **23**, 183–200.
Boschma, R.A. and Frenken, K. (2006), 'Why is economic geography not an evolutionary science? Towards an evolutionary economic geography', *Journal of Economic Geography*, **6** (3), 273–302.
Boschma, R. and K. Frenken (2009a), 'Some notes on institutions in evolutionary economic geography', *Economic Geography*, **85** (2), 151–8.
Boschma, R. and K. Frenken (2009b), 'Technological relatedness and regional branching', in H. Bathelt, M.P. Feldman and D.F. Kogler (eds), *Dynamic Geographies of Knowledge Creation and Innovation*, London: Routledge, Taylor and Francis, forthcoming.
Boschma, R.A. and S. Iammarino (2009), 'Related variety, trade linkages and regional growth in Italy', *Economic Geography*, **85** (3), 289–311.
Boschma, R.A. and J.G. Lambooy (1999), 'Evolutionary economics and economic geography', *Journal of Evolutionary Economics*, **9**, 411–29.
Boschma, R.A. and R. Martin (2007), 'Constructing an evolutionary economic geography', *Journal of Economic Geography*, **7** (5), 537–48.
Boschma, R.A. and A.L.J. Ter Wal (2007), 'Knowledge networks and innovative performance in an industrial district: the case of a footwear district in the South of Italy', *Industry and Innovation*, **14** (2), 177–99.
Boschma, R.A. and R. Wenting (2007), 'The spatial evolution of the British automobile industry. Does location matter?', *Industrial and Corporate Change*, **16** (2), 213–38.
Boschma, R., R. Eriksson and U. Lindgren (2009), 'How does labour mobility affect the performance of plants? The importance of relatedness and geographical proximity', *Journal of Economic Geography*, **9** (2), 169–90.
Boyer, R. (1988), 'Technical change and the theory of regulation', in G. Dosi, C. Freeman, R. Nelson, G. Silverberg and L. Soete (eds), *Technical Change and Economic Theory*, London: Pinter Publishers, pp. 67–94.
Brenner, T. (2004), *Local Industrial Clusters: Existence, Emergence and Evolution*, London/New York: Routledge.

Breschi, S. and F. Lissoni (2003), 'Mobility and social networks: localised knowledge spillovers revisited', CESPRI Working Paper 142, Bocconi University, Milan.

Breschi, S. and F. Malerba (1997), 'Sectoral innovation systems: technological regimes, Schumpeterian dynamics, and spatial boundaries', in C. Edquist (ed.), *Systems of Innovation. Technologies, Institutions and Organizations*, London/Washington: Pinter, pp. 130–56.

Breschi, S., F. Lissoni and F. Malerba (2003), 'Knowledge-relatedness in firm technological diversification', *Research Policy*, **32**, 69–87.

Broekel, T. and A. Meder (2008), 'The bright and dark side of cooperation for regional innovation performance', Papers in Evolutionary Economic Geography, No. 08.11, Utrecht: Department of Economic Geography, Utrecht University.

Brouwer, A.E. (2005), *Old Firms in the Netherlands. The Long-Term Spatial Implications of Firms' Identities and Embeddedness*, Utrecht/Groningen: KNAG/FRW, RUG, Netherlands Geographical Studies No. 329.

Buenstorf, G. and S. Klepper (2009), 'Heritage and agglomeration. The Akron tyre cluster revisited', *The Economic Journal*, **119**, 705–33.

Cantner, U. and H. Graf (2006), 'The network of innovators in Jena: an application of social network analysis', *Research Policy*, **35** (4), 463–80.

Cantner, U. and A. Meder (2007), 'Technological proximity and the choice of cooperation partner', *Journal of Economic Interaction and Coordination*, **2** (1), 45–65.

Cantwell, J.A. and S. Iammarino (2003), *Multinational Corporations and European Regional Systems of Innovation*, London: Routledge.

Cantwell, J. and G.D. Santangelo (2002), 'The new geography of corporate research in Information and Communications Technology (ICT)', *Journal of Evolutionary Economics*, **12**, 163–97.

Carlsson, B. (2003), 'Innovation systems. A survey of the literature from a Schumpeterian perspective', paper presented at the International Schumpeter Society Conference, Milano, June 2003.

Carroll, G.R. and M.T. Hannan (2000), *The Demography of Corporations and Industries*, Princeton, NJ: Princeton University Press.

Castellacci, F. (2006), 'A critical realist interpretation of evolutionary growth theorising', *Cambridge Journal of Economics*, **30** (6), 861–80.

Castells, M. (1996), *The Rise of the Network Society*, Oxford: Blackwell.

Chapman, K. (1991), *The International Petrochemical Industry. Evolution and Location*, Oxford: Basil Blackwell.

Cohen, W.M. and D.A. Levinthal (1990), 'Absorptive capacity: a new perspective on learning and innovation', *Administrative Science Quarterly*, **35**, 128–53.

Consoli, D. (2005), 'The dynamics of technological change in UK retail banking services. An evolutionary perspective', *Research Policy*, **34**, 461–80.

Cooke, P. (1992), 'Regional innovation systems. Competitive regulation in the new Europe', *Geoforum*, **23**, 365–82.

Cooke, P. (2001), 'Regional innovation systems, clusters, and the knowledge economy', *Industrial and Corporate Change*, **10** (4), 945–74.

Cooke, P., M.G. Uranga and G. Extebarria (1998), 'Regional innovation systems: an evolutionary perspective', *Environment and Planning A*, **30**, 1563–84.

David, P.A. (1985), 'The economics of QWERTY', *American Economic Review (Papers and Proceedings)*, **75**, 332–7.

David, P.A. (2005), 'Path dependence in economic processes: implications for policy analysis in dynamical systems contexts', in K. Dopfer (ed.), *The Evolutionary Foundations of Economics*, Cambridge: Cambridge University Press, pp. 151–94.

Dopfer, K. (ed.) (2004), *The Evolutionary Foundations of Economics*, Cambridge: Cambridge University Press.

Dopfer, K. and J. Potts (2004), 'Evolutionary realism: a new ontology for economics', *Journal of Economic Methodology*, **11** (2), 195–212.

Dosi, G., C. Freeman, R. Nelson, G. Silverberg and L. Soete (eds) (1988), *Technical Change and Economic Theory*, London: Pinter Publishers.

Dunning, J.H. (1994), 'Re-evaluating the benefits of foreign direct investment', *Transnational Corporations*, **3**, 23–52.

Economic Geography (2009), Special Issue on the Role of Institutions in Evolutionary Economic Geography (ed. G. Grabher), *Economic Geography*, **85** (2), 119–82.

Edquist, C. (ed.) (1997), *Systems of Innovation. Technologies, Institutions and Organizations*, London/Washington: Pinter.

Essletzbichler, J. (2007), 'Diversity, stability and regional growth in the United States, 1975–2002', in K. Frenken (ed.), *Applied Evolutionary Economics and Economic Geography*, Cheltenham, UK and Northampton, MA, USA: Edward Edgar, pp. 203–29.

Essletzbichler, J. (2009), 'Evolutionary economic geography, institutions and political economy', *Economic Geography*, **85** (2), 159–65.
Essletzbichler, J. and D.L. Rigby (2005), 'Competition, variety and the geography of technology evolution', *Tijdschrift voor Economische en Sociale Geografie*, **96** (1), 48–62.
Feldman, M.P. and D.B. Audretsch (1999), 'Innovation in cities. Science-based diversity, specialization and localized competition', *European Economic Review*, **43**, 409–29.
Fine, B. (2000), 'Endogenous growth theory: a critical assessment', *Cambridge Journal of Economics*, **24**, 245–65.
Foster, J. (1997), 'The analytical foundations of evolutionary economics: from biological analogy to economic self organization', *Structural Change and Economic Dynamics*, **8**, 427–51.
Foster, J. (2000), 'Competitive selection, self-organisation and Joseph A. Schumpeter', *Journal of Evolutionary Economics*, **10**, 311–28.
Foster, J. (2005), 'From simplistic to complex systems in economics', *Cambridge Journal of Economics*, **29**, 873–92.
Freeman, C. (1987), *Technology Policy and Economic Performance. Lessons from Japan*, London: Pinter.
Freeman, C. and C. Perez (1988), 'Structural crisis of adjustment, business cycles and investment behaviour', in G. Dosi, C. Freeman, R. Nelson, G. Silverberg and L. Soete (eds), *Technical Change and Economic Theory*, London: Pinter, pp. 38–66.
Frenken, K. (ed.) (2007), *Applied Evolutionary Economics and Economic Geography*, Cheltenham, UK and Northampton, MA, USA: Edward Elgar.
Frenken, K. and R.A. Boschma (2007), 'A theoretical framework for evolutionary economic geography: industrial dynamics and urban growth as a branching process', *Journal of Economic Geography*, **7** (5), 635–49.
Frenken, K., F.G. Van Oort and T. Verburg (2007), 'Related variety, unrelated variety and regional economic growth', *Regional Studies*, **41** (5), 685–97.
Gertler, M.S. (2009), 'Rules of the game: the place of institutions in regional economic change', paper presented at the Regional Studies Lecture at the AAG 2008 Conference, Boston.
Gilsing, V., B. Nooteboom, W. Vanhaverbeke, G. Duysters and A. van den Oord (2007), 'Network embeddedness and the exploration of novel technologies. Technological distance, betweenness centrality and density', *Research Policy*, **37**, 1717–31.
Giuliani, E. (2007), 'The selective nature of knowledge networks in clusters: evidence from the wine industry', *Journal of Economic Geography*, **7** (2), 139–68.
Giuliani, E. and M. Bell (2005), 'The micro-determinants of meso-level learning and innovation. Evidence from a Chilean wine cluster', *Research Policy*, **34**, 47–68.
Glaeser, E.L. (2005), 'Reinventing Boston: 1630–2003', *Journal of Economic Geography*, **5**, 119–53.
Glaeser, E.L., H.D. Kallal, J.A. Schinkmann and A. Shleifer (1992), 'Growth in cities', *Journal of Political Economy*, **100**,1126–52.
Glückler, J. (2007), 'Economic geography and the evolution of networks', *Journal of Economic Geography*, **7** (5), 619–34.
Grabher, G. (1993), 'The weakness of strong ties: the lock-in of regional development in the Ruhr area', in G. Grabher (ed.), *The Embedded Firm*, London: Routledge, pp. 255–77.
Grabher, G. (2009), 'Yet another turn? The evolutionary project in economic geography', *Economic Geography*, **85** (2), 119–27.
Grabher, G. and D. Stark (1997), 'Organizing diversity: evolutionary theory, network analysis and postsocialism', *Regional Studies*, **31**, 533–44.
Gunderson, L. and C.S. Holling (eds) (2002), *Panarchy: Understanding Transformations in Human and Natural Systems*, Washington, DC: Island Press.
Hagedoorn, J. (2002), 'Inter-firm R&D partnerships: an overview of major trends and patterns since 1960', *Research Policy*, **31** (4), 477–92.
Hall, P.G. and P. Preston (1988), *The Carrier Wave. New Information Technology and the Geography of Innovation 1846–2003*, London: Unwin Hyman.
Hannan, M.T., G.R. Carroll, E.A. Dutton and J.C. Torres (1995), 'Organizational evolution in a multinational context: entries of automobile manufacturers in Belgium, Britain, France, Germany and Italy', *American Sociological Review*, **60** (4), 509–28.
Hannan, M.T. and J.H. Freeman (1989), *Organizational Ecology*, Cambridge, MA: Harvard University Press.
Hassink, R. (2005), 'How to unlock regional economies from path dependency? From learning region to learning cluster', *European Planning Studies*, **13** (4), 521–35.
Hassink, R. and D.H. Shin (2005), 'Guest editorial: the restructuring of old industrial areas in Europe an Asia', *Environment and Planning A*, **37**, 571–80.
Hausmann, R. and B. Klinger (2007), 'The structure of the product space and the evolution of comparative

advantage', CID working paper No. 146, Center for International Development, Harvard University, Cambridge.

Heiner, R.A. (1983), 'The origin of predictable behaviour', *The American Economic Review*, **73**, 560–95.

Henderson, J.V., A. Kuncoro and M. Turner (1995), 'Industrial development in cities', *Journal of Political Economy*, **103**, 1067–85.

Hidalgo, C.A., B. Klinger, A.L. Barabasi and R. Hausmann (2007), 'The product space conditions the development of nations', *Science*, **317**, 482–7.

Hill, E., H. Wial and H. Wolman (2008), 'Exploring regional resilience', Working Paper WP2008–04, Institute of Urban and Regional development, University of California, Berkeley.

Hodgson, G. (1993), *Economics and Evolution: Bringing Life Back into Economics*, Cambridge: Polity Press.

Hodgson, G.M. (2009), 'Agency, institutions and Darwinism in evolutionary economic geography', *Economic Geography*, **85** (2), 167–73.

Hohenberg, P.M. and L.H. Lees (1995), *The Making of Urban Europe 1000–1994*, Cambridge, MA: Harvard University Press.

Hollingsworth, J.R. (2000), 'Doing institutional analysis: implications for the study of innovations', *Review of International Political Economy*, **7**, 595–644.

Iammarino, S. (2005), 'An evolutionary integrated view of regional systems of innovation. Concepts, measures and historical perspectives', *European Planning Studies*, **13**, (4), 497–519.

Iammarino, S. and P. McCann (2006), 'The structure and evolution of industrial clusters. Transactions, technology and knowledge spillovers', *Research Policy*, **35**, (7), 1018–1036.

Jacobs, J. (1969), *The Economy of Cities*, New York: Vintage Books.

Journal of Economic Geography (2007), Special Issue on Evolutionary Economic Geography (eds R. Boschma and R.L. Martin), *Journal of Economic Geography*, **7**, 537–672.

Jovanovic, M.N. (2009), *Evolutionary Economic Geography. Location of Production and the European Union*, London: Routledge.

Klaes, M. (2004), 'Evolutionary economics: in defence of "Vagueness"', *Journal of Economic Methodology*, **11**, 359–76.

Klepper, S. (1997), 'Industry life-cycles', *Industrial and Corporate Change*, **6**, 145–82.

Klepper, S. (2007), 'Disagreements, spinoffs, and the evolution of Detroit as the capital of the U.S. automobile industry', *Management Science*, **53**, 616–31.

Knoben, J. (2007), *A Spider and the Stickiness of its Web. The Causes and Consequences of Spatial Firm Mobility*, Tilburg: Tilburg University, unpublished PhD thesis.

Krugman, P.R. (1991), 'Increasing returns and economic geography', *Journal of Political Economy*, **99** (3), 483–99.

Lambooy, J.G. (1984), 'The regional ecology of technological change', in J.G. Lambooy (ed.), *New Spatial Dynamics and Economic Crisis*, Tampere: Finnpublishers, pp. 63–77.

Lambooy, J.G. (2002), 'Knowledge and urban economic development: An evolutionary perspective', *Urban Studies*, **39** (5–6), 1019–35.

Lambooy, J.G. and R.A. Boschma (2001), 'Evolutionary economics and regional policy', *Annals of Regional Science*, **35** (1), 113–33.

Lundvall, B.-Å. (ed.) (1992), *National Innovation Systems: Towards a Theory of Innovation and Interactive Learning*, London: Pinter.

McCann, P. (2007), 'Observational equivalence? Regional studies and regional science', *Regional Studies*, **41**, 1209–21.

MacKinnon, D., A. Cumbers, A. Pyke, K. Birch and R. McMaster (2009), 'Evolution in economic geography. Institutions, political economy and adaptation', *Economic Geography*, **85** (2), 129–50.

Maggioni, M.A. (2002), *Clustering Dynamics and the Location of High-Tech Firms*, Heidelberg/New York: Physica-Verlag.

Malerba, F. (2002), 'Sectoral systems of innovation and production', *Research Policy*, **31** (2), 247–64.

Malerba, F. (ed.) (2004), *Sectoral Systems of Innovation: Concepts, Issues and Analyses of Six Major Sectors in Europe*, Cambridge: Cambridge University Press.

Malmberg, A. and P. Maskell (1997), 'Towards an explanation of regional specialization and industry agglomeration', *European Planning Studies*, **5**, 25–41.

Malmberg, A. and Maskell, P. (2002), 'The elusive concept of localization economies: towards a knowledge-based theory of spatial clustering', *Environment and Planning A*, **34**, 429–49.

Markusen, A. (1985), *Profit Cycles, Oligopoly and Regional Development*, Cambridge, MA: MIT Press.

Markusen, A. (1999), 'Fuzzy concepts, scanty evidence, policy distance: the case for rigour and policy relevance in critical regional studies', *Regional Studies*, **33**, 869–84.

Marshall, M. (1987), *Long Waves of Regional Development*, London: Macmillan.

Martin, R. (1999), 'The new "geographical turn" in economics: some critical reflections', *Cambridge Journal of Economics*, **23** (1), 65–91.

Martin, R. (2000), 'Institutional approaches in economic geography', in E. Sheppard and T.J. Barnes (eds), *A Companion to Economic Geography*, Oxford and Malden MA: Blackwell Publishing, pp. 77–94.
Martin, R. (ed.) (2008), *Economic Geography: Critical Essays in Human Geography* (Contemporary Foundations of Space and Place Series), Aldershot: Ashgate.
Martin, R. (2010a), 'The "new economic geography": credible models of the economic landscape?', in A. Leyshon, R. Lee, L. McDowell and P. Sunley (eds), *The Sage Companion to Economic Geography*, London: Sage (in press).
Martin, R. (2010b), 'Rethinking regional path dependence: from lock-in to evolution', *Economic Geography*, **86** (1), 1–27.
Martin, R. and P.J. Sunley (2001), 'Rethinking the "Economic" in economic geography: broadening our vision or losing our focus', *Antipode*, **33** (2), 148–61.
Martin, R. and P. Sunley (2003), 'Deconstructing clusters: chaotic concept or policy panacea?', *Journal of Economic Geography*, **3** (1), 5–35.
Martin, R. and P. Sunley (2006), 'Path dependence and regional economic evolution', *Journal of Economic Geography*, **6** (4), 395–437.
Martin, R. and P. Sunley (2007a), 'Complexity thinking and evolutionary economic geography', *Journal of Economic Geography*, **7**, 573–601.
Martin, R. and P. Sunley (eds) (2007b), *Economic Geography* (Critical Concepts in the Social Sciences Series), London: Routledge.
Maskell, P. (2001a), 'The firm in economic geography', *Economic Geography*, **77** (4), 329–44.
Maskell, P. (2001b), 'Towards a knowledge-based theory of the geographical cluster', *Industrial and Corporate Change*, **10**, 919–41.
Maskell, P. and A. Malmberg (2007), 'Myopia, knowledge development and cluster evolution', *Journal of Economic Geography*, **7** (5), 603–18.
Menzel, M.P. (2008), *Dynamic Proximities. Changing Relations by Creating and Bridging Distances*, Papers in Evolutionary Economic Geography, No. 08.16, Utrecht: Department of Economic Geography, Utrecht University.
Menzel, M.P. and D. Fornahl (2009), 'Cluster life cycles. Dimensions and rationales of cluster evolution', *Industrial and Corporate Change*, Advance Access, published 22 July, doi: 10.1093/icc/dtp036.
Metcalfe, S. (1994), 'The economic foundations of technology policy: equilibrium and evolutionary perspectives', in M. Dodgson and R. Rothwell (eds), *The Handbook of Industrial Innovation*, Cheltenham, UK and Northampton, MA, USA: Edward Elgar, pp. 409–512.
Metcalfe, J.S. (1998), *Evolutionary Economics and Creative Destruction*, London: Routledge.
Metcalfe, J.S. (2005), 'Systems failure and the case for innovation policy', in P. Llerena and M. Matt (eds), *Innovation Policy in a Knowledge Based Economy*, Berlin: Springer.
Metcalfe, J.S. and J. Foster (eds) (2004), *Evolution and Economic Complexity*, Cheltenham, UK and Northampton, MA, USA: Edward Elgar.
Metcalfe, J.S., J. Foster and R. Ramlogan (2006), 'Adaptive economic growth', *Cambridge Journal of Economics*, **30**, 7–32.
Mokyr, J. (2005), 'Long-term economic growth and the history of technology', in P. Aghion and S. Durlauf (eds), *Handbook of Economic Growth*, Amsterdam: Elsevier, pp. 1113–80.
Morgan, K. (1997), 'The learning region: institutions, innovation and regional renewal', *Regional Studies*, **31**, (5), 491–503.
Morrison, A. (2008), 'Gatekeepers of knowledge within industrial districts: who they are, how they interact', *Regional Studies*, **42**, 817–35.
Murmann, J.P. (2003), *Knowledge and Competitive Advantage. The Co-evolution of Firms, Technology, and National Institutions*, Cambridge: Cambridge University Press.
Neffke, F. (2009), 'Productive places. The influence of technological change and relatedness on agglomeration externalities', PhD thesis, Utrecht University, Utrecht.
Neffke, F. and M. Svensson Henning (2009), *Revealed Relatedness: Mapping Industry Space*, Papers in Evolutionary Economic Geography, Utrecht: University of Utrecht.
Nelson, R.R. (ed.) (1993), *National Innovation Systems. A Comparative Analysis*, Oxford & New York: Oxford University Press.
Nelson, R.R. (1995), 'Co-evolution of industry structure, technology and supporting institutions, and the making of comparative advantage', *International Journal of the Economics of Business*, **2** (2), 171–84.
Nelson, R.R. (2002), 'Bringing institutions into evolutionary growth theory', *Journal of Evolutionary Economics*, **12**, 17–28.
Nelson, R.R. and S.G. Winter (1977), 'In search of a useful theory of innovation', *Research Policy*, **6**, 36–76.
Nelson, R.R. and S.G. Winter (1982), *An Evolutionary Theory of Economic Change*, Cambridge, MA and London: Belknap Press.

Nooteboom, B. (2000), *Learning and Innovation in Organizations and Economies*, Oxford: Oxford University Press.
Nooteboom, B. and E. Stam (2008), *Micro-Foundations for Innovation Policy*, Amsterdam/Chicago: Amsterdam University Press/Chicago University Press.
Nooteboom, B., W. Van Haverbeke, G. Duysters, V. Gilsing and A. van den Oord (2007), 'Optimal cognitive distance and absorptive capacity', *Research Policy*, **36**, 1016–34.
Norton, R.D. (1979), *City Life-Cycles and American Urban Policy*, New York: Academic Press.
Norton, R.D. and J. Rees (1979), 'The product cycle and the spatial decentralization of American manufacturing', *Regional Studies*, **13**, 141–51.
Ormerod, P. (2005), *Why Most Things Fail. Evolution, Extinction and Economics*, New York: Pantheon Books.
Orsenigo, L., F. Pammoli, M. Riccaboni, A. Bonnaccorsi and G. Turchetti (1998), 'The evolution of knowledge and the dynamics of an industry network', *Journal of Management and Governance*, **1**, 147–75.
Otto, A. and S. Kohler (2008), 'The contribution of new and young firms to the economic development of clusters in Germany. Comparative analysis of a growing, a mature and a declining cluster', in U. Blien and G. Maier (eds), *The Economics of Regional Clusters. Networks, Technology and Policy*, Cheltenham, UK and Northampton, MA, USA: Edward Elgar, pp. 167–84.
Ozman, M. (2009), 'Inter-firm networks and innovation: a survey of literature', *Economics of Innovation and New Technology*, **18** (1), 39–67.
Pelikan, P. and G. Wegner (eds) (2003), *The Evolutionary Analysis of Economic Policy*, Cheltenham, UK and Northampton, MA, USA: Edward Elgar.
Peng, M.W. (1995), 'Foreign direct investment in the innovation-driven stage. Towards a learning option perspective', in M.B. Green and R.B. McNaughton (eds), *Location of Foreign Direct Investment. Geographic and Business Approaches*, Aldershot: Ashgate, pp. 29–43.
Penrose, E. (1959), *The Theory of the Growth of the Firm*, Oxford: Oxford University Press.
Potts, J. (2000), *The New Evolutionary Micro-economics: Complexity, Competence and Adaptive Behaviour*, Cheltenham, UK and Northampton, MA, USA: Edward Elgar.
Powell, W.W., K. Koput and L. Smith-Doerr (1996), 'Interorganizational collaboration and the locus of innovation: networks of learning in biotechnology', *Administrative Science Quarterly*, **41**, 116–45.
Powell, W.W., D.R. White, K.W. Koput and J. Owen-Smith (2005), 'Network dynamics and field evolution: the growth of interorganizational collaboration in the life sciences', *American Journal of Sociology*, **110** (4), 1132–205.
Pred, A.R. (1967), *Behavior and Location. Foundations for a Geographic and Dynamic Location Theory*, Lund: Lund Studies in Geography 27.
Pumain, D. and F. Moriconi-Ebrard (1997), 'City size distributions and metropolisation', *Geojournal*, **43** (4), 307–14.
Ramlogan, R. and J.S. Metcalfe (2006), 'Restless capitalism: a complexity perspective on modern capitalist economies', in E. Garnsey and J. McGlade (eds), *Complexity and Co-evolution: Continuity and Change in Socio-Economics Systems*, Cheltenham, UK and Northampton, MA, USA: Edward Elgar, pp. 115–46.
Raspe, O. and F.G. van Oort (2006), 'The knowledge economy and urban economic growth', *European Planning Studies*, **14**, 1209–34.
Rigby, D.L. and J. Essletzbichler (1997), 'Evolution, process variety, and regional trajectories of technological change in US manufacturing', *Economic Geography*, **73** (3), 269–84.
Rosenberg, N. (1969), 'The direction of technical change: inducement mechanisms and focusing devices', *Economic Development and Cultural Change*, **18**, 1–24.
Rosenberg, N. (1976), *Perspectives on Technology*, Cambridge: Cambridge University Press.
Rosser, J.B. (ed.) (2009), *Handbook of Research on Complexity*, Cheltenham, UK and Northampton, MA, USA: Edward Elgar.
Saxenian, A.L. (2006), *The New Argonauts. Regional Advantage in a Global Economy*, Cambridge, MA: Harvard University Press.
Schamp. E.W. (2005), 'Decline of the district, renewal of firms: an evolutionary approach to footwear production in the Pirmasens area, Germany', *Environment and Planning A*, **37**, 617–34.
Scott, A.J. (1988), *New Industrial Spaces. Flexible Production Organization and Regional Development in North America and Western Europe*, London: Pion.
Scott, A.J. and M. Storper (1987), 'High technology industry and regional development. A theoretical critique and reconstruction', *International Social Science Journal*, **112**, 215–32.
Setterfield, M. (1997), *Rapid Growth and Relative Decline. Modelling Macroeconomic Dynamics with Hysteresis*, London: Macmillan.
Shapira, P. and G. Fuchs (eds) (2005), *Rethinking Regional Innovation and Change: Path Dependency or Regional Breakthrough*, New York: Springer.
Sheppard, E. and T.J. Barnes (eds) (2000), *A Companion to Economic Geography*, Oxford and Malden, MA: Blackwell Publishing.

Simmie, J. (2005), 'Innovation and space: a critical review of the literature', *Regional Studies*, **39** (6), 789–804.
Simmie, J. and J. Carpenter (eds) (2007), *Path Dependence and the Evolution of City Regional Development*, Working Paper Series No. 197, Oxford: Oxford Brookes University.
Simon, H.A. (1955), 'On a class of skew distribution functions', *Biometrika*, **42** (3–4), 425–40.
Staber, U. (1997), 'An ecological perspective on entrepreneurship in industrial districts', *Entrepreneurship and Regional Development*, **9**, 45–64.
Staber, U. (2001), 'The structure of networks in industrial districts', *International Journal of Urban and Regional Research*, **25**, 537–52.
Staber, U. (2007), 'A matter of distrust: explaining the persistence of dysfunctional beliefs in regional clusters', *Growth and Change*, **38**, 341–63.
Stam, E. (2007), 'Why butterflies don't leave: locational behavior of entrepreneurial firms', *Economic Geography*, **83**, 27–50.
Sternberg, R. and O. Arndt (2001), 'The firm or the region: what determines the innovation behavior of European firms?', *Economic Geography*, **77** (4), 364–82.
Storper, M. (1992), 'The limits to globalization. Technology districts and international trade', *Economic Geography*, **68** (1), 60–93.
Storper, M. (1997), *The Regional World*, New York: Guilford Press.
Storper, M. and R. Walker (1989), *The Capitalist Imperative. Territory, Technology and Industrial Growth*, New York: Basil Blackwell.
Sunley, P. (2008), 'Relational economic geography: a partial understanding or a new paradigm?', *Economic Geography*, **84** (1), 1–26.
Swann, P. and M. Prevezer (1996), 'A comparison of the dynamics of industrial clustering in computing and biotechnology', *Research Policy*, **25**, 1139–57.
Taylor, M. (1986), 'The product life cycle model. A critique', *Environment and Planning A*, **18**, 751–61.
Taylor, M.J. and B.T. Asheim (2001), 'The concept of the firm in economic geography', *Economic Geography*, **77** (4), 315–28.
Taylor, P.J. and R. Aranya (2008), 'A global "urban roller coaster"? Connectivity changes in the world city network, 2000–2004', *Regional Studies*, **42** (10), 1–16.
Taylor, P.J., B. Derudder, P. Saey and F. Witlox (eds) (2007), *Cities in Globalization. Theories, Policies, Practices*, London: Routledge.
Ter Wal, A.L.J. (2009), *The Spatial Dynamics of the Inventor Network in German Biotechnology. Geographical Proximity versus Triadic Closure*, Utrecht: Department of Economic Geography.
Ter Wal, A.L.J. and R.A. Boschma (2009), 'Co-evolution of firms, industries and networks in space', *Regional Studies*, in press.
Tödtling, F. and M. Trippl (2005), 'One size fits all? Towards a differentiated regional innovation policy approach', *Research Policy*, **34**, 1203–19.
Torre, A. and A. Rallet (2005), 'Proximity and localization', *Regional Studies*, **39** (1), 47–60.
Uzzi, B. (1996), 'The sources and consequences of embeddedness for the economic performance of organizations: the network effect', *American Sociological Review*, **61**, 674–98.
van Duijn, J.J., and J.G. Lambooy (1982), *Technological Innovation and Regional Economic Growth: A Meso-economic Analysis,* Amsterdam: Faculty of Economics Research Memorandum, No. 8207.
Van Geenhuizen, M. and P. Nijkamp (2006), 'Learning regions in an evolutionary context. Policymaking for high technology firms', *International Journal of Entrepreneurship and Innovation Management*, **6** (3), 265–82.
Veblen, T. (1898), 'Why is economics not an evolutionary science?', *Quarterly Review of Economics*, **12**, 373–97.
Visser, E.J. (2009), 'The complementary dynamic effects of clusters and networks', *Industry and Innovation*, **16** (2), 167–95.
Wall, R. (2009), *Netscape. Cities and Global Corporate Networks*, PhD manuscript, Rotterdam: Erasmus University Rotterdam.
Wenting, R. (2008), 'Spinoff dynamics and the spatial formation of the fashion design industry, 1858–2005', *Journal of Economic Geography*, **8**, 593–614.
Wenting, R. and K. Frenken (2008), 'Firm entry and institutional lock-in: an organizational ecology analysis of the global fashion design industry', *Papers in Evolutionary Economic Geography*, 0801, available at: http://econ.geo.uu.nl/peeg/peeg.html
Wezel, F.C. (2005), 'Location-dependence and industry evolution. Founding rates in the United Kingdom motorcycle industry, 1895–1993', *Organization Studies*, **26** (5), 729–54.
Wimmer, A. and R. Kössler (eds) (2006), *Understanding Change: Models, Methodologies and Metaphors*, Basingstoke: Palgrave Macmillan.
Winter, S.G. and G. Szulanski (2001), 'Replication as strategy', *Organization Science*, **12**, 730–43.
Wintjens, R. (2001), *Regional Economic Effects of Foreign Companies. A Research on the Embeddedness of*

American and Japanese Companies in Dutch Regions (in Dutch), Netherlands Geographical Studies No. 286, Utrecht: KNAG/University Utrecht.
Witt, U. (2003), *The Evolving Economy: Essays on the Evolutionary Approach to Economics*, Cheltenham, UK and Northampton, MA, USA: Edward Elgar.
Witt, U. (2006), *Evolutionary Economics*, Papers on Economics and Evolution, No. 0605, Jena: Max Planck Institute of Economics, Evolutionary Economics Group.

PART 1

CONCEPTUAL CHALLENGES IN EVOLUTIONARY ECONOMIC GEOGRAPHY

2 Generalized Darwinism and evolutionary economic geography

Jürgen Essletzbichler and David L. Rigby

1. Introduction

Over the last few years evolutionary metaphors and concepts have become increasingly popular with economic geographers (Amin, 1999; Barnes, 1997; Boschma and Frenken, 2006; Boschma and Lambooy, 1999; Boschma and Van der Knaap, 1997; Essletzbichler, 1999; Essletzbichler and Rigby, 2005a, 2005b, 2007; Frenken, 2007; Frenken and Boschma, 2007; Grabher, 1993; Grabher and Stark, 1997; Hudson, 2001; Martin, 2000; Martin and Sunley, 2006, 2007; Rigby and Essletzbichler, 1997, 2006; Storper, 1997; Webber et al., 1992). For the most part, however, concepts such as path-dependence and lock-in have been deployed in a rather ad hoc fashion and the development of a broader evolutionary approach to economic geography has not progressed very far.

Three major points of departure in the theoretical development of evolutionary economics are readily identified in economic geography, including complexity theory (Frenken, 2006; Martin and Sunley, 2007; Plummer and Sheppard, 2006), path- dependence (Martin and Sunley, 2006) and generalized Darwinism (Essletzbichler and Rigby, 2007; Rigby and Essletzbichler, 1997). Each of these approaches emphasizes different moments of the evolutionary process. Complexity theory focuses on the creation of variety, path-dependence stresses the retention of existing information and knowledge, and generalized Darwinism examines how a population of heterogeneous entities evolve through interaction among themselves and with the environment that they help shape.

In this chapter we outline a general approach to economic geography based on generalized Darwinism. We focus on this approach because we believe that it is consistent with, yet more general than, approaches based on complexity or path-dependence. In the following section we discuss the basic principles of evolutionary economics. We then explore an evolutionary model of economic dynamics where economic agents are located in different geographical spaces. We seek to show how competition between those agents, based on the core evolutionary principles of variety, selection and continuity, may produce distinct economic regions sharing properties that differentiate them from elsewhere. The competitive dynamics of agents *in* regions are then linked to the dynamics *of* regions.

2. Evolutionary economics: basic principles and concepts

Modern evolutionary economic theory emerged in the 1970s largely in opposition to the core assumptions of mainstream, neoclassical economics (Nelson and Winter, 1974). Evolutionary economics targeted issues that neoclassical theory seemed least capable of explaining in depth, namely economic growth (Nelson, 1995; Nelson and Winter, 1982; Verspagen, 2001), technological change (Arthur, 1983, 1989, 1994; David, 1985; Dosi, 1982; Dosi et al., 1988; Pavitt, 2005), industrial evolution (Klepper, 2001; Klepper and

Graddy, 1990; Utterback and Abernathy, 1975), the nature of competition and the role of institutions and routines in guiding individual behavior (Hodgson, 1988; Nelson, 2001; Veblen, 1898). As the approach matured, debates over appropriate foundations and the style of evolutionary economic theory emerged (Hodgson, 2002; Knudsen, 2002; Witt, 2004). This section briefly reviews these debates and leads to an outline of the fundamental arguments of an evolutionary approach to economic change.

Evolutionary theory emerged from at least three theoretical positions: non-linear dynamics and complexity theory; path-dependence; and evolutionary biology. The more recent engagement with evolutionary economics can be traced to the pioneering efforts of Nelson and Winter (1982). In part, this work originated in attempts to engage the selection arguments of Alchian (1950) and to rebuff the 'as if' claims of Friedman (1953), the search to ground marginalist claims of profit maximization within an evolutionary selection framework (Vromen, 1995). However, Nelson and Winter's work became much more, offering a model of evolutionary economic dynamics, and standing as a prelude to the most sustained period of development that evolutionary economics has yet seen.

While most evolutionary economists are united by rejection of the assumptions of full information and perfect rationality that underpin mainstream neoclassical economics, they remain far from formulating a common research paradigm, agreement on basic principles, or even on the best way to carry the framework forward (Hanappi, 1995). Hodgson (1993, 2002), Knudsen (2002, 2004) and Witt (2004) discuss the different ways that evolutionary arguments have been extended into the domain of economic activity.

A first strategy involves application of the neo-Darwinian theory of natural selection directly to human behavior, and is closely linked to socio-biology (Becker, 1976; Wilson, 1975). Because humans are products of natural selection, and because economic actions result from human activity, neo-Darwinians believe that economic behavior should be explicable on the basis of how it is correlated with the genetic fitness of individuals (Witt, 2004; p. 127). Because this approach only applies to genetically determined forms of behavior, because it ignores human intentionality and demands complete separation of genetic information, or social equivalents, from the selection environment, it has tended to gather little recent support (Hodgson, 2004; Vromen, 2004). A second strategy is to make use of Darwinian principles in a purely heuristic or metaphorical fashion. In this case, different concepts are imported to derive 'metaphoric inspiration' or to construct analogies between evolutionary biology and evolutionary economics (see Essletzbichler and Rigby, 2007; Hodgson, 1993). This approach is widely applied in economics and economic geography, though perhaps a little too uncritically. While there is little question of the usefulness of sharing ideas across the natural and social sciences, there are concerns with the inconsistency of eclecticism (Fincher, 1983), and with the too rigid interpretation of analogies across disciplinary boundaries. Witt (2004) also believes certain concepts that are clearly defined in evolutionary biology are impossible to define in social systems and thus a close or true analogy is unlikely. The most promising strategy for development of evolutionary economics and its extension into economic geography is based on generalized Darwinism. Generalized Darwinism asserts that the core principles of evolution provide a general theoretical framework for understanding evolutionary change in complex population systems (from physical to social systems), but that the meaning of those principles and the way that they operate are specific to each domain (Hodgson, 2002, 2005; Hodgson and Knudsen, 2006; Knudsen, 2004). Thus, to understand economic evolution from the approach of generalized Darwinism demands understanding what the key concepts of variation, selection and continuity might represent in the economy, how those concepts are put into motion, or embedded within a dynamic system of economic competition, and how they are influenced by other mechanisms specific to that system. As Vromen (2004) notes, the aim of this approach is not to see whether economic evolution can fit the general schema of generalized Darwinism, but, rather, to see whether analysis of economic dynamics using these principles provides novel insight into the movement of the economy over time.

The fundamental aim of evolutionary economics is thus to understand the dynamic processes that jointly influence the behavior of firms and the market environment in which they operate (Nelson and Winter, 1982). These processes can be considered evolutionary in the sense that the capitalist economy consists of competing agents that differ in at least some characteristics (heterogeneity) that influence individual prospects for economic growth (selection), and that change more or less slowly over time (heredity), both shaping and being shaped by the environment within which future competition unfolds (Hodgson, 1993, 2002; Metcalfe, 1998).

Within the capitalist mode of production, firms differ from one another across a series of dimensions – product type, technology, organizational form, location and the behavioral routines adopted to regulate processes of investment, labor management, technological search, and so on. This heterogeneity is an inevitable byproduct of competition and innovation within an economy where production is carried out by private firms motivated by profit but limited by information asymmetries and uneven capabilities (Alchian, 1950). Uncertain of the future, firms control whatever they can as best they can, seeking competitive advantage by increasing the efficiency of production. For most, however, efficiency is unknown until they enter the market. In this competitive environment firms are compelled to innovate, to search for new products and develop new markets, to experiment with new sources of inputs, new processes of production and organizational routines, sure only in the knowledge that others are doing the same (Schumpeter, 1942). Indeed, it is this constant imperative to innovate that sustains economic variety and provides the energy that fuels evolutionary change.

The competitive process of market selection regulates the profitability of individual firms, their prospects for growth, and their ability to generate new routines. Selection thereby alters the environment within which future decisions are made: it pushes some firms out of the market, encourages others to enter, and reshuffles the relative efficiency of competing agents. It is important to note that the process of selection does not necessarily identify and reward more efficient firms. Rather, sales are distributed unevenly across competitors in a market. On average, those firms that produce a commodity more efficiently are better able to translate revenues into profits and thus, given a certain propensity to invest, increase their relative size at the expense of firms that are relatively inefficient (Metcalfe, 1998). Some firms deliberately attempt to alter the selection environment in which they find themselves, perhaps by differentiating the commodity they offer for sale and thus competing in a particular niche market. Regardless, they are still hostage to the same uncertainty that pervades all unregulated markets. Notice that selection does not result in 'progressive evolution', that is, the survival of the fittest, most adaptive or most efficient. Fitness or efficiency is always context dependent and defined relative to a locally given environment and not according to some global maximum.

What is efficient in one context may be inefficient in another. Furthermore, the technical definition of selection (Price, 1970, 1995) does not necessarily entail improvement in overall fitness or efficiency (Andersen, 2004a, 2004b; Frank, 1998; Hodgson and Knudsen, 2006).

For selection to operate, a certain level of stability, or inertia, in firm characteristics is required. In a world of infinite malleability and instantaneous adaptability variety disappears and there is nothing to select. Sunk capital investments, contractual relationships, organizational strategies and the accumulated knowledge base of the firm generate the necessary institutional inertia that allows selection to operate. Nelson and Winter (1982) argue that firms develop routines to cope with decision-making and that these change only slowly in reaction to shifts in the environment. It is these behavioral routines that play the role of heredity in the evolutionary economic model of the firm, preserving some continuity over time in firm characteristics (Nelson and Winter, 1982). While they tend to be relatively stable in the short run, the behavioral routines of firms do change, the result of profit-induced search, learning, imitation, and chance, by adaptation to the changing economic environment and as part of their efforts to strategically manipulate that environment. As long as economic agents are boundedly rational and as long as some form of inertia exists that enables differences in behavior to persist, forces of selection will operate and economic evolution will run its course (Hodgson, 2001; Hodgson and Knudsen, 2004; Simon, 1957).

These broad principles of evolutionary change have found increasing application within economics and related fields over the past 20 years or so. Following Nelson and Winter (1974, 1982), a prominent focus of this literature has been industrial dynamics (Baldwin, 1995; Klepper, 1996, 2001). In support of these arguments, Nelson (1995), Cohen and Levinthal (1989), Rigby and Essletzbichler (2000, 2006) explore the existence and persistence of variety in firm technologies and organizational routines. Saviotti (1996), Cantner and Hanusch (2001) and Essletzbichler and Rigby (2005b) propose different notions of variety including product and process variety. Search routines and patterns of innovation that lead to the generation of heterogeneity are explored by David (1975), Metcalfe and Gibbons (1986), Dosi (1988, 1997) and Webber et al. (1992). As new forms of establishment-level data have been developed, so analysis of industry and technological evolution has increasingly taken advantage of these sources (Audretsch, 1995; Baldwin, 1995; Bartelsman and Doms, 2000; Rigby and Essletzbichler, 2006). At the same time, newer models of evolutionary dynamics have emerged, building on more sophisticated understandings of population dynamics (Anderson, 2004a, 2004b; Iwai, 1984a, 1984b; Metcalfe, 1998).

Evolutionary arguments appear to hold much promise for the analysis of economic dynamics. Nonetheless, a number of potential stumbling blocks remain:

1. For some, evolution does not require a theory of innovation. For evolution to proceed it is sufficient (and necessary) merely that variety exists (Metcalfe, 1998). However, innovation, or the creation of heterogeneity, is a necessary process of long-run evolution in systems that are not degenerate. Within the capitalist economic system, it is impossible to comprehend the process of competition without innovation and the shadow of uncertainty that it casts. Generalized Darwinism handles the creation of variety through the process of generative selection (Hodgson

and Knudsen, 2006), but changes occurring through self-transformation or development will have to be included as well (Foster and Metcalfe, 2001).

2. The principle of self-organization, i.e. the emergence of patterns through interaction between entities that cannot be reduced to properties of those entities, holds much potential for explaining evolutionary processes not only in social and cultural but also in biological systems (Depew and Weber, 1995, 1996; Foster, 1997, 2000, 2001; Kauffman, 1993). However, it is unclear whether the principle of self-organization can be accommodated within the framework of generalized Darwinism, although a synthesis of the two frameworks appears possible and even necessary (Depew and Weber, 1995, 1996; Foster and Metcalfe, 2001; Hodgson and Knudsen, 2006; Kauffman, 1993; Metcalfe, 2005). Self-organization can explain how order may emerge from interacting agents 'but itself it explains neither (a) the characteristics of the agents that interact to create the emergent order, (b) how the emergent order reacts to competing social orders, nor (c) more generally how an emergent order adapts and survives in the broader social and natural environment' (Hodgson and Knudsen, 2006; p. 9). The principles of self-organization and selection appear thus complementary rather than contradictory.
3. Within the evolutionary framework more squarely, it is still unclear whether firms are the most appropriate unit of selection within the economy.[1] Further, we do not really know the characteristics of those units that are most critical in terms of selection, nor how the pressure of selection shifts as firms and industries mature.
4. Additional work is also required to understand the interaction between individual units in populations of interest and the 'environment' that they shape and within which they evolve. For example, what evolutionary processes give rise to path dependence, why do some emergent properties of systems get 'locked-in' for shorter or longer periods, and through precisely what mechanisms is such stability maintained and how is it overturned?

3. Generalized Darwinism and evolutionary economic geography

While economic geographers frequently employ evolutionary concepts and metaphors such as routines, path-dependence, lock-in, and co-evolution in a descriptive manner (Barnes, 1997), recent work attempts to ground an evolutionary economic geography on more solid theoretical foundations (Essletzbichler and Rigby, 2007; Frenken, 2006; Frenken and Boschma, 2007; Martin and Sunley, 2006, 2007). In this section of the chapter we outline the conceptual foundations of an evolutionary approach to regional uneven development. Our aim is to reveal the utility of the core principles of generalized Darwinism within the domain of economic geography. We do not insist that evolutionary arguments in general and those based on an abstract model of generalized Darwinism are necessarily the most useful for the study of regional economic dynamics, merely that they provide a different perspective that offers novel insights. We do believe that an evolutionary framework has considerably more to offer than a few descriptive metaphors. In the limited space available, we cannot hope to provide a fully fledged evolutionary model of economic dynamics in space. Rather, we seek to provide a general introduction to the role of variety, selection and continuity in shaping economic dynamics, the way that these processes are influenced by space, and in turn influence the evolution of regional economies.

The basic concepts of an evolutionary understanding of economic change were outlined above. These form the essential building blocks around which evolutionary accounts of regional economic dynamics have to be structured. Those accounts rest on a population of competing entities of which at least some have unique characteristics that lead to a differential allocation of resources that constrain behavior. The movement of the system as a whole reflects changes in the relative weights of the different entities, the birth of new competitors and elimination of existing ones, and processes of transformation that alter the characteristics of individual units. As we move to consider the evolution of the space economy, we must also examine the ways that selection environments (spaces of competition) are produced and transformed by the actions of individual economic agents, broader coalitions and institutions, and how the characteristics of those spaces influence patterns of economic change.

Evolutionary approaches to economic dynamics have a number of potential points of departure. The units of selection that ground an evolutionary account could be firms, workers, specific technologies, competing modes of regulation, or the routines and institutions found in particular places. We have tended to choose the business establishment (plant) as our basic unit of analysis and so privilege the economic dynamics that originate in plant-level competition. A consideration of evolutionary dynamics over space raises the question of whether regions themselves might be considered as units of selection. We suggest that they should not, as this might suggest that regions act as homogeneous wholes which would represent another form of spatial fetishism. However, we do accept the hierarchical view of evolution endorsed by a growing number of researchers and believe that location influences the behavior and fitness of individual entities within a region (Gould and Vrba, 1982; Gowdy, 1992; Hodgson, 2001; Levins and Lewontin, 1985; Lewontin, 1970; Vrba and Gould, 1986).

While not a unit of selection, perhaps it makes more sense to think of regions (at different scales) as forming selection environments within which and across which evolutionary processes operate. Individual businesses compete within one or more such selection environments, facing different pressures in each, and more or less able to modify those environments either directly or indirectly. Regional characteristics, or the forms of spatial selection environments, evolve through time, all shaped by local histories of political-economic development and some influenced by the patterns and character of development elsewhere. The same broad forces of competition that undergird the birth and death of individual business units, the production and application of new technologies and organizational forms, also create and destroy the political and economic environments in which competition unfolds: technologies, institutions and the firms and regions in which they are embedded are ephemeral. Regions then are not simply passive containers that bound the activities of political and economic agents, they represent ecologies that are continually in flux, the boundaries and forms of which are continuously contested. And, as spatial and institutional environments change, the competitive pressures acting on political-economic agents shift, inducing yet further rounds of search and innovation. Evolutionary economic geographies must then focus on evolution *in* a region as well as the evolution *of* regions. We develop these concepts through the arguments below.

Evolution in the region

We start our analysis of regional economic change by considering the evolution of a population of plants that compete with one another within a common selection environment, perhaps understood as a single region. The general delineation of this region might be the result of technology, fixed over the short run by the costs of transport, from political fiat, or from development of an institutional fabric given local coherence by shared history. Over this short run, economic growth and change in the economy may be understood as a simple aggregate of changes in the characteristics of individual business units. Over the longer run, organizational forms and institutional relationships may shift, altering the structure of competition and generating more profound shifts in the regional political economy. Formal analysis of population dynamics results from the work of Fisher (1930) and his fundamental theorem of natural selection that states the rate of increase of aggregate fitness in a population is dependent on the variance of fitness across the units that comprise that population. In our context, we might restate Fisher's theorem as aggregate regional economic performance depends on the variance in levels of performance across the plants within the region. It is important to note that Fisher's argument captures a simple distance-from-mean replicator dynamics, where, on average, more (less) efficient plants experience faster (slower) growth rates. Plant entry and exit can be incorporated within this model of selection.

Fisher's theorem is important for understanding population thinking and the process of selection. It leads to a consideration of the properties that are selected for within a population of interest and it focuses attention on the strength of the process of selection. However, Fisher's theorem alone offers only a partial account of evolutionary dynamics because it does not explain the generation of novelty. Within the economy, at least, there is surely little question that competitive advantage and growth hinge on technological change, on processes of learning, imitation and innovation, the deliberate search by economic agents to improve their performance (Andersen, 2004a, 2004b; Metcalfe, 1998; Price, 1970). Price extends Fisher's arguments to show how the effects of selection and technological change can be integrated to account for shifts in aggregate performance across a population of competing economic units. Technological change here is used broadly to represent any alteration in production technology, organizational form, behavioral routines or related characteristics of plants that impact efficiency.

Figure 2.1 illustrates these ideas, providing a stylized representation of the population of business establishments (plants) that define, in part, a region's economy. Here we ignore the formation and history of institutional arrangements and other characteristics that more fully specify the coalescence of forces and relations of production that comprise the political-economy of a region at some time. Figure 2.1 describes plants in terms of their technology (by a vector of capital and labor input coefficients; see Essletzbichler and Rigby, 2005a). Thus, the symbols contained within the ellipse at time t represent all the production units in operation within the region at that time. The different symbols identify business establishments that vary in terms of size and whether or not they remain in business from time t to time $t + 1$. The location of the symbols fixes the particular combination of capital and labor coefficients, the technology, used by each plant. Thus, at time t, the solid circles represent plants of different size that remain in business at times t and $t + 1$. We label these plants as incumbents. The crosses represent plants that are in business at time t, but exit the region's economy by time $t + 1$. These plants are known as exits. The aggregate technology of the region at time t is the output weighted average of the techniques of all plants operating in that time. Similarly, at time $t + 1$, the aggregate technology of the region is the output weighted average of plant technologies in use

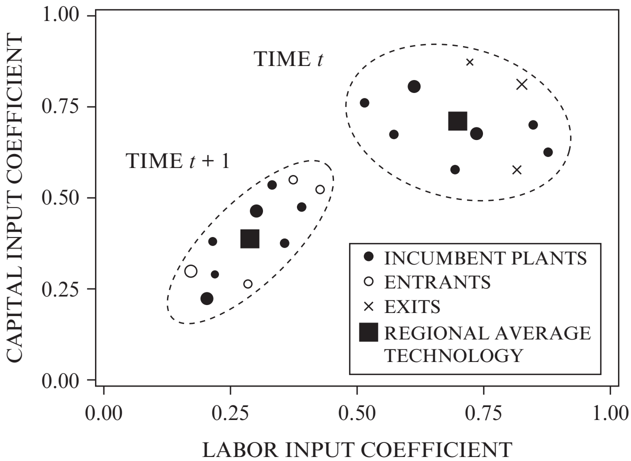


Figure 2.1 Plant technology, variety and aggregate regional change

at that time. The population of plants operating at time $t + 1$ is the set of incumbents from time t plus new plants that were not in business at time t. These new plants we call entrants and they are denoted by open circles. In Figure 2.1, the ellipses that bound the techniques of production for a given time provide a representation, and a measure, of the extent of technological heterogeneity in the regional economy. The existence of variety is key to the operation of evolutionary mechanisms of change.

We can use the population-based arguments of evolutionary accounts to outline the different processes responsible for shifts in aggregate regional technology over time. Until the recent development of plant-level micro-data, these processes were hidden in aggregate measures of technology, the movements of which were often, and wrongly, ascribed to technological change in general. Indeed, changes in aggregate regional techniques of production result from the following processes, that operate in quite distinct ways at different times and in different places (Rigby and Essletzbichler, 2006):

1. Technological change in incumbent plants.
2. Differential growth (selection) of incumbent plants.
3. Plant exit.
4. Plant entry.

Thus, in Figure 2.1, as incumbent plants adopt new technologies between times t and $t + 1$, so they move in technology space and tend to pull the region's average technology along with them. The adoption of different techniques of production in incumbent plants might represent innovation, the introduction of new techniques to the economy, or imitation, the copying of techniques already in use. Even if incumbent plants do not alter their techniques of production, as long as they experience rates of growth that differ from one another, aggregate technology in the region will shift as individual plants exert more or less weight on the aggregate measure. The entry and exit of plants over time also shifts the region's technology, so long as the entering and exiting plants have techniques of production that are different from the regional average. In general, exiting plants tend to be less efficient than average, while entrants tend to have techniques that are more widely dispersed.

Separating the influence of these processes on aggregate measures of technology can make use of a variety of decompositions. The most general of these, avoiding covariance terms, is outlined here. Thus, the change in the labor coefficient in region *r*, between time *t* and time *t* + 1 is:

$$l^{r\,t+1} - l^{r\,t} = \Delta l^r = \sum_{p \in I} \bar{s}_p^r \Delta l_p^r + \sum_{p \in I} (\bar{l}_p^r - \bar{l}^r)\,\Delta s_p^r + \sum_{p \in N} s_p^{r\,t+1}(l_p^{r\,t+1} - \bar{l}^r) - \sum_{p \in X} s_p^{r\,t}(l_p^{r\,t} - \bar{l}^r) \qquad (2.1)$$

In equation (2.1), *s* represents market share and is defined at the plant level as $s_p^{rt} = x_p^{rt}/\Sigma_p x_p^{rt}$, and x_p represents the output of plant *p*. In turn, Δ represents the change in the value of the following variable between years *t* and *t* + 1 and the ‾ notation represents the average value of the variable over the years *t* and *t* + 1. The summation subscript *I* denotes the set of incumbent plants, the summation subscript *N* represents entrants, and the summation subscript *X* represents exiting plants. In turn, the terms on the right hand side of equation (2.1) capture the following influences on aggregate technology: technological change in incumbent plants; changes in the market share of incumbent plants; plant entry and plant exit. Changes in the aggregate capital coefficient can be decomposed in exactly the same way.

The theoretical frameworks developed by Fisher and Price help us understand the broad movement of the economy from a population perspective (Mayr, 1984). Building on these foundations, more detailed arguments outline how processes of competition and cooperation give rise to the particular configurations of technologies and institutions that we observe. In terms of technology, considerable attention has been given to the way in which heterogeneity shapes both the direction and pace of change. From the work of Habakkuk (1962), David (1975), Metcalfe and Gibbons (1986) and Webber et al. (1992), we know that the form or shape of technological variety influences the direction of imitation within an economy, and how innovation is guided by relative prices. Those prices are part of the selection environment within which plants compete, they are generated by the choices of individual plants, and through market processes of supply and demand, as well as through political contest, particularly in the case of wages.

The extent of variety within a selection environment is also thought to control the pace of aggregate change, after Fisher (1930). Much less clear, however, are the precise linkages between aggregate change, the strength of selection and the generation and destruction of heterogeneity. In particular, how does the variance in plant characteristics influence the pressure of selection; to what extent can inefficiency (perhaps thought of as a measure of variety) be subsidized as a hedge against the lock-in of characteristics that may prove unprofitable in the long run; how does lock-in occur and how is it overturned? Here there is great need for careful historical accounts of industrial and regional development, of the generation of new products and processes of production, of competition between firms and technologies and resulting aggregate dynamics. Ideally, that history would trace the strategies and fortunes of individual business units, linking economic data for establishments and industries in particular places with firm ethnographies and other narratives to understand the development of place-specific institutions and the broader evolution of the selection environment.

It is increasingly clear that plants and industries do not evolve in a vacuum but co-evolve with other economic agents and alongside the broader institutional settings that sometimes develop within the regions in which they are embedded (Freeman, 1995; Nelson, 1995). Setterfield (1993) discusses how the rules and behavioral norms that comprise institutions emerge from the interaction of economic agents and the structures that regulate their activity, in a process of hysteresis. David (1992) offers a similar claim, with positive feedbacks generating that hysteresis. These arguments are consistent with the view that institutions are endogenously generated among populations of actors that engage in sustained social interaction. Indeed, Setterfield (1993) envisions a process of institutional creation and selection that is explicitly evolutionary in nature. If this interaction is bounded by the region that comprises the selection environment, it is likely that a set of broad regional social structures will emerge to coordinate economic activities. These institutional arrangements exist outside the boundaries of individual economic agents to deal with common problems in a manner consistent with the claims of Granovetter (1985).

Many questions surround the evolution of institutions at the level of the plant and among groups of agents at the level of the industry and region. While some routines might be developed through deliberate search processes, others come about by trial and error. Cumulative feedback mechanisms between the existing institutional environment and local agents may lead to path-dependence and lock-in of particular institutions. As with technologies, institutions are neither predictable ex ante nor do they necessarily represent the most efficient institutional form possible. Different institutional frameworks tend to operate at different scales of analysis. In some instances (e.g. specific industries) local specificity may matter for explaining evolutionary change, while the local scale may be irrelevant in others or overshadowed by the effects of global or national institutions. The actual geographic extent over which institutions exert influence will be context specific.

Evolution of regions

When we shift attention away from the single region to consider competition among plants located in different regions, some interacting and some not, then place-specific characteristics become increasingly important to the performance of individual plants and to the regions in which they operate. And, once efficiency criteria become defined across regions, when previously independent selection environments merge, for example, then a new evolutionary dynamic develops that couples evolution *in* regions with the evolution *of* regions. In evolutionary terms, this leads to models that combine intra-population dynamics with inter-population dynamics. It is to these issues that we now turn.

Where social and geographic distances between regions are relatively large, transport costs sufficiently high and connectivity between places thus relatively low, or institutional differences sufficiently strong, regional economies may evolve independently. Economic change within each region can be reduced to an analysis of population dynamics, and in a uniform selection environment, those dynamics are controlled largely by heterogeneity in the characteristics of individual agents. Of greater interest in this

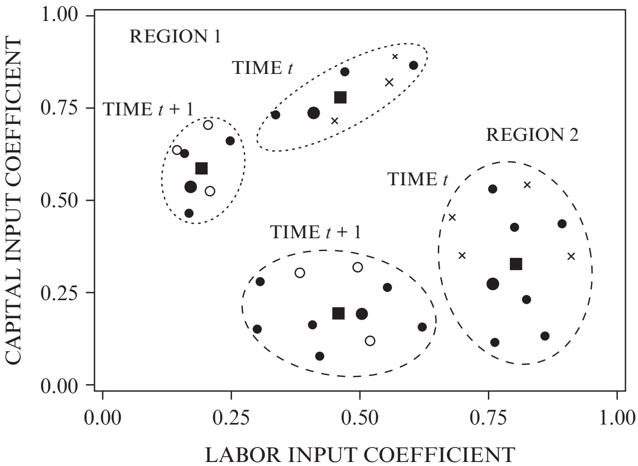


Figure 2.2 Regional evolution along different technological trajectories

multi-region setting is the character of evolution in different places. Regions co-evolve with the individual agents located in them. Over time economic agents located in different regions may develop distinct characteristics and routines, and regions themselves may become increasingly differentiated in terms of the mix of industries, the form and depth of institutions, organizational forms and other place-bound resources. Figure 2.2 illustrates this situation, showing two regions with different technologies moving along different trajectories.

Saxenian (1994) and Gertler (2005) provide clear evidence of institutional variation between regions. Evidence of technological variations between regions is also clear. Learning processes, search and knowledge flows all tend to be highly localized (Arrow, 1962, Jaffe et al., 1993, Lundvall, 1988). This reflects familiarity with an existing knowledge base, technological interdependencies and network relations (Gertler, 2005). Localization is also reinforced by sharply declining returns to investment in R&D efforts that are distant from existing technology and by steep distance decay effects as well as sharp discontinuities around the boundaries of specific knowledge bases. The geographical constraints on technology development suggest that technologies will evolve along relatively distinct pathways as the plants and firms of different regions follow innovatory trajectories conditioned by their history and geography (Clark, 1985; Dosi, 1982; Sahal, 1981). Empirical evidence strongly confirms these claims (David, 1975; Essletzbichler and Rigby, 2005b; Habakkuk, 1962; Rigby and Essletzbichler, 1997, 2006).

If we conceive regions as selection environments, as spatial units that comprise economic agents (plants, workers, consumers) and the relations between those agents (social networks, input–output linkages, untraded interdependencies), all embedded within institutional environments (governance systems, business cultures) at least some of which are locally circumscribed, then it is likely that the economic performance of regions will vary because of differences in their characteristics. The characteristics of individual economic agents, of collections of agents, and of regions in general that are most important in determining economic success are largely unknown, and are likely to vary across industries and regions and over time. For these reasons, predicting the fortunes of individual regions is difficult.

Regions themselves are not static entities but evolve over time as the endogenous population of economic agents and their characteristics shift, and as those agents deliberately try to influence the organizational and institutional environment within which they operate. That environment might be radically transformed by the development of new technologies that alter the friction of distance, or as institutional controls on borders and trade relax, redrawing the boundaries of the selection environment, perhaps exposing businesses to one another that formerly competed in separate spaces. In this way, regions and institutional regimes themselves might be considered as emergent properties of the process of capitalist competition. Indeed, as characteristic bundles of agents and place-specific attributes, regions increasingly become central to the strategies of economic agents, as spaces of contestation with resources to control, or when conditions dictate as spaces that can be remapped, shifting the boundaries of competition in particular directions.

Figure 2.3 depicts a situation where at time t two regions with distinct selection environments (here shown in technology space) merge at time $t + 1$ into one new region that forms a single selection environment within which all plants compete. Such change has been created by political shifts, most clearly perhaps in the case of the reunification of Germany (see Grabher and Stark, 1997), but it is more commonly the result of transport and communications improvements that have led to formerly independent economies being exposed to one another through trade. Consolidation of trade relations often heralds price adjustments, a remapping of the competitive standing of individual businesses, institutional conflict, shifts in markets and consequent changes in industrial scale and structure as capital is moved between sectors and spaces to exploit new opportunities for profit.

Selection environments, as spaces of competition, evolve through time along with the population of political and economic agents that help shape them. Selection envi-

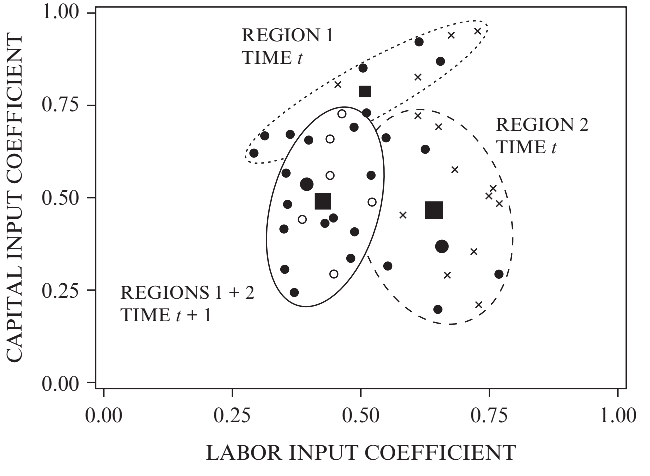


Figure 2.3 Merger of formerly independent selection environments

ronments might evolve gradually, and in relatively predictable ways, as institutional relationships and regulations are developed and applied, and as technology moves along established trajectories. Those environments might also change rapidly and in less predictable ways, perhaps as a result of radical technological innovation, radical policy shifts, or following the opening of trade and competition between formerly autarkic regions. Significant change in the characteristics of a selection environment may give rise to a thoroughgoing rearrangement of the efficiency of individual business units and to more or less intense selective pressure that heralds further change in the nature of that environment. Such changes will have a clear spatial expression if the remapping of efficiency favors the economic agents and institutions of one region over another. These points should make clear that efficiency is relative and always defined in the context of a selection environment that is produced, at least in part, by the actions of political and economic agents.

The evolutionary arguments above have considerably enriched our understanding of the regional dimensions of competition. Regions, like plants and firms, may be characterized by their variety in terms of technological, organizational, and institutional characteristics. And, as in the economy more generally, technological progress and economic change at the regional level are shaped by the competitive processes that create and destroy variation and that select certain techniques, institutions and organizational forms over others. While these competitive pressures manifest themselves in various product markets and in the differential performance of plants, it would be wrong to view the performance of regions as nothing more than the performance of the business units that they contain. Stripped of their association with the social and institutional fabric that defines the familiar political-economy of a home space/local selection environment, even the most competitive plants might wither. With or without common histories of practice, once populations of plants become isolated, over time they will tend to develop different characteristics and exhibit increasingly distinct histories of development. These histories will shape the selection environment in which those plants are located, producing and reproducing regions with characteristics that emerge from the interplay of the forces of capitalist competition as well as from other social pressures.

It is clear today, as production becomes increasingly fragmented, that firms are paying more attention than ever to the characteristics of particular places as they search for more attractive sites of accumulation. This fragmentation of production is, at the same time, integrating regions across the world economy. This integration implies significant changes in the selection environments within which plants compete. Regions are no longer simple repositories of independent plants, other economic agents and local institutional forces that can be interpreted as the containers within which competition works itself out. Rather, regions might be more accurately conceived as evolving bundles of attributes, some place-specific, others exotic, reflecting the inconstant population of economic agents in the region and the routines they have acquired through intra-firm and inter-firm networks that span multiple spaces. In this more complex environment, the processes of selection, of variety creation and destruction still function, though it is a much more difficult task to show how they can be geographically isolated to account for the uneven development of regions. Indeed, individual plants now appear to be competing across a hierarchy of relatively unstable selection environments that span local, national and even global spaces. In some respects, perhaps, we might conceive of firms and other agents as creating their own 'regional' selection environments through firm-specific divisions of labor across sectors and across spaces, acting also to promote particular institutional regimes over others and thus attempting to tilt the balance of power in their favor.

4. Conclusion

An evolutionary model of change offers a theoretically rich framework for analysis within economic geography. Like most theoretical models, evolutionary accounts highlight specific aspects of the real world at the expense of others. At a general level, evolutionary approaches to the economy appeal to us for the following reasons. First, models of evolution pertain to open, complex systems that may include socio-economic systems. Within these systems the future is unpredictable and, hence, the economy neither gravitates towards some predetermined equilibrium state nor progresses 'naturally' to some higher state. Second, the notion of equilibrium, a system at rest, is antithetical to an evolutionary model of economic change where uncertainty generates continuous experimentation and search for advantage. An evolutionary approach to economic geography thus focuses on dynamics, on the processes that create and destroy spatial assemblies of political-economic activity, the institutional relationships that shape much of that activity and its geographical heterogeneity. Third, evolutionary explanations are based on a population approach that celebrates the diverse characteristics and behaviors of individual agents and shows how macroeconomic order can emerge from the seemingly chaotic actions of myriad competitors. That order does not have to be generated by appeals to perfect information and rationality, as in the core neoclassical arguments that underpin general equilibrium. Fourth, an evolutionary approach is concerned with the behaviors of individual agents and their inter-relationships within environments that both constrain, and are shaped by, those agents. Because the dynamics of evolutionary economic change are linked to the domains of activity within which political-economic agents operate, the actions of those agents may only be understood through their location in both historical and spatial dimensions. Fifth, it is important to recognize that the actions of individual agents occur within contexts that are shaped by broader institutional structures that are themselves created and that evolve over time. Some of these institutions are more durable than others, such as the capitalist mode of production. A critical element of an evolutionary economic geography is understanding how and where these institutions are created, and how they are maintained.

We identified a number of different approaches within evolutionary economics and argued that generalized Darwinism has most to offer economic geography. Generalized Darwinism rests on the key evolutionary processes of variety, selection and continuity. The way that these processes operate within the domain of problems typically examined by economic geographers must be developed within that field of study. While we do not claim that metaphorical inspiration from evolutionary biology is completely unhelpful, we do claim that careful development of the core principles of generalized Darwinism within economic geography helps shed light on key agents and institutions of change, on the characteristics of those agents, on the characteristics of the natural and social environments within which competition operates and on which of those characteristics exerts most leverage across the different processes that shape differential growth and uneven spatial development. We go on to argue that while a number of researchers within economic geography have used the concepts of path-dependence, lock-in and selection in their work, such application is insufficient to render the analysis evolutionary. These same concepts can be developed within non-evolutionary frameworks, given appropriate assumptions. An example would be the possibility of multiple equilibria in new economic geography models. Assuming increasing returns and transportation costs in the manufacturing sector, the level of transportation costs will determine whether we observe a core–periphery structure or whether all regions diversify their economies. Because there are a range of transportation costs where both core–periphery and diversified patterns are possible, we get the possibility of multiple equilibria and a role for history or policy to influence the outcome. In this sense, the actual (equilibrium) outcome may depend on specific historical conditions such as the geographical distribution of the population prior to changes in the transportation costs (Fujita et al., 1999). The actually observed outcome may thus be sensitive to historical conditions.

Based on the general principles outlined above, an evolutionary economic geography is broadly concerned with analysis of the impact of plant location on the creation of variety, on the transmission of information from one period to the next, and on the extent to which the selection environment shapes spaces of competition and cooperation. Economic geographers may be interested in how location influences the pace and direction of the generation of novelty. Why are plants in some regions more innovative than others? Why do plants in different regions specialize in the creation of different products and technologies and organize their businesses differently? Why do institutions differ across space and how do they influence the creation of novelty? Work on regional innovation systems, relational assets and regional institutions is relevant here. The existing literature contributes to a detailed understanding of those processes but focuses only on one moment of the evolutionary process. A second moment of evolutionary analysis must also explain the survival of entities with certain characteristics over others. This problem entails linking the characteristics of agents such as firms or institutions with the broader environment in which they operate. While markets for products may be global (entailing a global selection environment), regional institutional differences (national tax policies or regional innovation policies) may raise or dampen firm-level efficiency and so influence firm success in global markets. A third evolutionary moment focuses on retention. Because evolution requires a certain level of stability in the characteristics of agents and the selection environments within which they operate, we have to understand how information is transmitted over time. Firm routines are generally identified as carriers of information in firms, and institutions are often identified as carriers of 'collective knowledge' in regions and countries. Institutional approaches within economic geography are critical to understand the role of institutions as carriers of information, how institutions are created and how they evolve over time.

Note

1. Multi-level selection theory or group selection are possible ways out of this dilemma and potentially important for the development of an evolutionary economic geography (Sober and Wilson, 1998; Vromen, 2001). Group selection refers to the idea that certain characteristics can emerge in groups of individual entities because of the benefits they confer to groups regardless of those characteristics' effects on the fitness of individual entities within the group. For instance, a group of firms in one region may develop strong altruistic behavior because cooperation has resulted in knowledge spillovers and other externalities that raised their efficiency levels relative to firms in another region (but not relative to other firms within the region)

where strong competition failed to produce those externalities. In this case selection across regions would favor the group of altruists over the group of competitors although individual altruists transplanted to the 'competitive region' would have no chance of survival. Multi-level selection models would attempt to isolate the relative importance of group versus individual (within-group) selection on changes in aggregate population characteristics.

References

Alchian, A. (1950), 'Uncertainty, evolution and economic theory', *Journal of Political Economy*, **58**, 211–22.

Amin, A. (1999), 'An institutionalist perspective on regional economic development', *International Journal of Urban and Regional Research*, **23**, 365–78.

Andersen, E. (2004a), 'Population thinking, Price's equation and the analysis of economic evolution', Paper submitted to *Evolutionary and Institutional Economics Review*.

Andersen, E. (2004b), 'From Schumpeter's failed econometrics to modern evometric analysis: creative destruction as a tale of two effects', Paper presented at the Conference of the International Schumpeter Society, Milan, 9–12 June.

Arrow, K. (1962), 'The economic implications of learning by doing', *Review of Economic Studies*, **29**, 22–43.

Arthur, W.B. (1983), 'On competing technologies and historical events small events: the dynamics of choice under increasing returns', Working Paper WP-83–90, International Institute for Applied Systems Analysis (IIASA), Vienna.

Arthur, W.B. (1989), 'Competing technologies, increasing returns, and lock-in by historical events', *Economic Journal*, **99**, 116–31.

Arthur, W.B. (1994), *Increasing Returns and Path Dependence in the Economy*, Ann Arbor: University of Michigan Press.

Audretsch, D.B. (1995), *Innovation and Industry Evolution*, Cambridge, MA: MIT Press.

Baldwin, J.R. (1995), *The Dynamics of Industrial Competition*, Cambridge: Cambridge University Press.

Barnes, T. (1997), 'Theories of accumulation and regulation: bringing back life to economic geography', in R. Lee and J. Wills (eds), *Geographies of Economies*, London: Arnold, pp. 231–24.

Bartelsman, E. and M. Doms (2000), 'Understanding productivity: lessons from longitudinal microdata', *Journal of Economic Literature*, **38**, 569–94.

Becker, G. (1976), *The Economic Approach to Human Behavior*, Chicago: University of Chicago Press.

Boschma, R. and K. Frenken (2006), 'Why is evolutionary economic geography not an evolutionary science?', *Journal of Economic Geography*, **6**, 273–302.

Boschma, R. and J. Lambooy (1999), 'Evolutionary economics and economic geography', *Journal of Evolutionary Economics*, **9**, 411–29.

Boschma, R. and G.A. Van der Knaap (1997), 'New technology and windows of locational opportunity: indeterminacy, creativity and chance', in J. Reijnders (ed), *Economics and Evolution*, Cheltenham, UK and Lyme, MA, USA: Edward Elgar, pp. 171–202.

Cantner, U. and H. Hanusch (2001), 'Heterogeneity and evolutionary change: empirical conception, findings and unresolved issues', in J. Foster and J.S. Metcalfe (eds), *New Frontiers in Evolutionary Economics*, Cheltenham, UK and Northampton, MA, USA: Edward Elgar, pp. 228–69.

Clark, K.B. (1985), 'The interaction of design hierarchies and market concepts in technological evolution', *Research Policy*, **14**, 235–51.

Cohen, W.M. and D.A. Levinthal (1989), 'Innovation and learning: the two faces of R&D', *Economic Journal*, **99**, 569–96.

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

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

David, P. (1992), 'Heroes, herds and hysteresis in technological history: Thomas Edison and "The Battle of Systems" reconsidered', *Industrial and Corporate Change*, **1**, 129–80.

Depew, D.J. and B.H. Weber (1995), *Darwinism Evolving*, Cambridge, MA and London: MIT Press.

Depew, D.J. and B.H. Weber (1996), 'Natural selection and self-organization', *Biology and Philosophy*, **11**, 33–65.

Dosi, G. (1982), 'Technological paradigms and technological trajectories: a suggested interpretation of the determinants and directions of technical change', *Research Policy*, **11**, 147–62.

Dosi, G. (1988), 'Sources, procedures and microeconomic effects of innovation', *Journal of Economic Literature*, **26**, 1120–71.

Dosi, G. (1997), 'Opportunities, incentives and the collective patterns of technological change', *Economic Journal*, **107**, 1530–47.

Dosi, G., C. Freeman, R. Nelson, G. Silberberg and L. Soete (eds) (1988), *Technical Change and Economic Theory,* London and New York: Pinter.
Essletzbichler, J. (1999), *Regional Differences in Technology Evolution in US Manufacturing*, Unpublished PhD dissertation, University of California Los Angeles.
Essletzbichler, J. and D. Rigby (2005a), 'Technological evolution as creative destruction of process heterogeneity. Evidence from US plant level data', *Economic Systems Research*, **17**, 25–45.
Essletzbichler, J. and D. Rigby (2005b), 'Competition, variety and the geography of technology evolution', *Tijdschrift voor Economische en Sociale Geografie*, **96**, 48–62.
Essletzbichler, J. and D.L. Rigby (2007), 'Exploring evolutionary economic geographies', *Journal of Economic Geography*, **7**, 549–71.
Fincher, R. (1983), 'The inconsistency of eclecticism', *Environment and Planning A*, **15**, 607–22.
Fisher, R.A. (1930), *The Genetical Theory of Natural Selection*, Oxford: Clarendon.
Foster, J. (1997), 'The analytical foundations of evolutionary economics: from biological analogy to economic self-organization', *Structural Change and Economic Dynamics*, **8**, 427–51.
Foster, J. (2000), 'Competitive selection, self-organization and Joseph A. Schumpeter', *Journal of Evolutionary Economics*, **10**, 311–28.
Foster, J. (2001), 'Competition, competitive selection and economic evolution', in P. Garrouste and S. Ioannides (eds), *Evolution and Path Dependence in Economic Ideas*, Cheltenham, UK and Northampton, MA, USA: Edward Elgar, pp. 107–32.
Foster, J. and S. Metcalfe (2001), 'Modern evolutionary economic perspectives: an overview', in J. Foster and J.S. Metcalfe (eds), *Frontiers of Evolutionary Economics*, Cheltenham, UK and Northampton, MA, USA: Edward Elgar, pp. 1–16.
Frank, S.A. (1998), *Foundations of Social Evolution*, Princeton, NJ: Princeton University Press.
Freeman, C. (1995), 'The "national system of innovation" in historical perspective', *Cambridge Journal of Economics*, **19**, 5–24.
Frenken, K. (2006), *Innovation, Evolution and Complexity Theory*, Cheltenham, UK and Northampton, MA, USA: Edward Elgar.
Frenken, K. (ed.) (2007), *Applied Evolutionary Economics and Economic Geography*, Cheltenham, UK and Northampton, MA, USA: Edward Elgar.
Frenken, K. and R. Boschma (2007), 'A theoretical framework for evolutionary economic geography: industrial dynamics and urban growth as a branching process', *Journal of Economic Geography*, **7**, 635–49.
Friedman, M. (1953), *Essays in Positive Economics*, Chicago: University of Chicago Press.
Fujita, M., P. Krugman and A.J. Venables (1999), *The Spatial Economy*, Cambridge, MA and London: MIT Press.
Gertler, M. (2005), *Manufacturing Culture: The Governance of Industrial Practice*, Oxford: Oxford University Press.
Gould, S.J. and E.S. Vrba (1982), 'Exaptation – a missing term in the science of form', *Paleobiology*, **8**, 4–15.
Gowdy, J.M. (1992), 'Higher selection processes in evolutionary economic change', *Journal of Evolutionary Economics*, **2**, 1–16.
Grabher, G. (1993), 'The weakness of strong ties – the lock-in of regional development in the Ruhr area', in G. Grabher (ed.), *The Embedded Firm*, London: Routledge, pp. 255–77.
Grabher, G. and D. Stark (1997), 'Organizing diversity: evolutionary theory, network analysis and postsocialism', *Regional Studies*, **31**, 533–44.
Granovetter, M. (1985), 'Economic action and social structure: the problem of embeddedness', *American Journal of Sociology*, **91**, 481–510.
Habakkuk, H.J. (1962), *American and British Technology in the 19th Century*, Cambridge: Cambridge University Press.
Hanappi, G. (1995), *Evolutionary Economics*. *The Evolutionary Revolution in the Social Sciences*, Aldershot: Avebury.
Hodgson, G. (1988), *Economics and Institutions*, Cambridge: Polity Press.
Hodgson, G. (1993), *Economics and Evolution. Bringing Life Back into Economics*, Oxford: Polity Press.
Hodgson, G. (2001), *How Economics Forgot History*. *The Problem of Historical Specificity in Social Science*, London and New York: Routledge.
Hodgson, G. (2002), 'Darwinism in economics: from analogy to ontology', *Journal of Evolutionary Economics*, **12**, 259–81.
Hodgson, G. (2004), 'Darwinism, causality and the social sciences', *Journal of Economic Methodology*, **11**, 175–94.
Hodgson, G. (2005), 'Generalizing Darwinism to social evolution: some early attempts', *Journal of Economic Issues*, **39**, 899–914.
Hodgson, G. and T. Knudsen (2004), 'The firm as an interactor: firms as vehicles for habits and routines', *Journal of Evolutionary Economics*, **14**, 281–307.

Hodgson, G. and T. Knudsen (2006), 'Why we need a generalized Darwinism, and why generalized Darwinism is not enough', *Journal of Economic Behavior and Organization*, **61**, 1–19.
Hudson, R. (2001), *Producing Places*, New York: Guilford Press.
Iwai, K. (1984a), 'Schumpeterian dynamics. An evolutionary model of innovation and imitation', *Journal of Economic Behavior and Organization*, **5**, 159–90.
Iwai, K. (1984b), 'Schumpeterian dynamics, part II: technological progress, firm growth and economic selection', *Journal of Economic Behavior and Organization*, **5**, 321–51.
Jaffe, A.B., M. Trajtenberg and R. Henderson (1993), 'Geographic localization of knowledge spillovers as evidenced by patent citations', *Quarterly Journal of Economics*, **108**, 577–98.
Kauffman, S. (1993), *The Origins of Order: Self-Organization and Selection in Evolution*, New York: Oxford University Press.
Klepper, S. (1996), 'Entry, exit, growth, and innovation over the product life cycle', *American Economic Review*, **86**, 562–83.
Klepper, S. (2001), 'The capabilities of new firms and the evolution of the U.S. automobile industry', *Industrial and Corporate Change*, **11**, 645–66.
Klepper, S. and E. Graddy (1990), 'The evolution of new industries and the determinants of market structure', *Rand Journal of Economics*, **21**, 27–44.
Knudsen, T. (2002), 'Economic selection theory', *Journal of Evolutionary Economics*, **12**, 443–70.
Knudsen, T. (2004), 'General selection theory and economic evolution: the price equation and the replicator/interactor distinction', *Journal of Economic Methodology*, **11**, 147–73.
Levins, R. and R. Lewontin (1985), *The Dialectical Biologist*, Cambridge, MA: Cambridge University Press.
Lewontin, R.C. (1970), 'The units of selection', *Annual Review of Ecology and Systematics*, 1, 114.
Lundvall, B.A. (1988), 'Innovation as an interactive process: from user–producer interaction to the national system of innovation', in G. Dosi, C. Freeman, R. Nelson, G. Silberberg and L. Soete (eds), *Technical Change and Economic Theory*, London and New York: Pinter, pp. 349–69.
Martin, R. (2000), 'Institutional approaches in economic geography', in E. Sheppard and T. Barnes (eds), *A Companion to Economic Geography*, Oxford: Blackwell, pp. 77–94.
Martin, R. and P. Sunley (2006), 'Path dependence and regional economic evolution', *Journal of Economic Geography*, **6**, 395–437.
Martin, R. and P. Sunley (2007), 'Complexity thinking and evolutionary economic geography', *Journal of Economic Geography*, **7**, 573–601.
Mayr, E. (1984), 'Typological versus population thinking', in E. Sober (ed.), *Conceptual Issues in Evolutionary Biology: An Anthology*, Cambridge, MA: MIT Press.
Metcalfe, S. (1998), *Evolutionary Economics and Creative Destruction*, London and New York: Routledge.
Metcalfe, S. (2005), 'Evolutionary concepts in relation to evolutionary economics', in K. Dopfer (ed.), *The Evolutionary Foundations of Economics*, Cambridge: Cambridge University Press, pp. 391–430.
Metcalfe, S. and M. Gibbons (1986), 'Technological variety and the process of competition', *Economie Appliquée*, **39**, 493–520.
Nelson, R. (1995), 'Recent evolutionary theorizing about economic change', *Journal of Economic Literature*, **33**, 48–90.
Nelson, R. (2001), 'The coevolution of technology and institutions as the driver of economic growth', in J. Foster and S. Metcalfe (eds), *Frontiers in Evolutionary Economics*, Cheltenham, UK and Northampton, MA, USA: Edward Elgar, pp. 19–30.
Nelson, R.R. and S. Winter (1974), 'Neoclassical and evolutionary theories of growth: critique and prospectus', *Economic Journal*, **84**, 886–905.
Nelson, R. and S. Winter (1982), *An Evolutionary Theory of Economic Growth*, Cambridge, MA: Harvard University Press.
Pavitt, K. (2005), 'Innovation processes', in J. Fagerberg, D.C. Mowery and R. Nelson (eds), *The Oxford Handbook of Innovation*, Chapter 4, Oxford: Oxford University Press.
Plummer, P. and E. Sheppard (2006), 'Geography matters: agency, structures and dynamics at the intersection of economics and geography', *Journal of Economic Geography*, **6**, 619–38.
Price, G.R. (1970), 'Selection and covariance', *Nature*, **227**, 520–21.
Price, G.R. (1995), 'The nature of selection', *Journal of Theoretical Biology*, **175**, 389–96.
Rigby, D. and J. Essletzbichler (1997), 'Evolution, process variety, and regional trajectories of technological change in U.S. manufacturing', *Economic Geography*, **73**, 269–84.
Rigby, D. and J. Essletzbichler (2000), 'Impacts of industry mix, technological change, selection and plant entry/exit on regional productivity growth', *Regional Studies*, **34**, 333–42.
Rigby, D. and J. Essletzbichler (2006), 'Technological variety, technological change and a geography of production techniques', *Journal of Economic Geography*, **6**, 45–70.
Sahal, D. (1981), *Patterns of Technological Innovation*, Reading, MA: Addison Wesley.

Saviotti, P. (1996), *Technological Evolution, Variety and the Economy*, Cheltenham, UK and Northampton, MA, USA: Edward Elgar.
Saxenian, A. (1994), *Regional Advantage*, Cambridge, MA: Harvard University Press.
Schumpeter, J. (1942), *Capitalism, Socialism and Democracy*, New York: Harper and Row.
Setterfield, M. (1993), 'A model of institutional hysteresis', *Journal of Economic Issues*, **27**, 755–74.
Simon, H. (1957), *Models of Man*, New York: John Wiley.
Sober, E. and D.S. Wilson (1998), *Unto Others: The Evolution and Psychology of Unselfish Behavior*, Cambridge, MA: Harvard University Press.
Storper, M. (1997), *The Regional World*, New York: Guilford Press.
Utterback, J.M. and W.J. Abernathy (1975), 'A dynamic model of product and process innovation', *Omega*, **3**, 639–56.
Veblen, T. (1898), 'Why is economics not an evolutionary science?', *Quarterly Journal of Economics*, **12**, 373–97.
Verspagen, B. (2001), 'Economic growth and technological change: an evolutionary interpretation', STI Working Papers 2001/1, OECD, Paris.
Vrba, E.S. and S.J. Gould (1986), 'The hierarchical expansion of sorting and selection: sorting and selection cannot be equated', *Paleobiology*, **10**, 217–28.
Vromen, J. (1995), *Economic Evolution: An Enquiry in the Foundations of New Institutional Economics*, London: Routledge.
Vromen, J. (2001), 'The human agent in evolutionary economics', in J. Laurent and J. Nightingale (eds), *Darwinism and Evolutionary Economics*, Cheltenham, UK and Northampton, MA, USA: Edward Elgar, pp. 184–208.
Vromen, J. (2004), 'Conjectural revisionary economic ontology: outline of an ambitious research agenda for evolutionary economics', *Journal of Economic Methodology*, **11**, 213–47.
Webber, M., E. Sheppard and D. Rigby (1992), 'Forms of technical change', *Environment and Planning*, A, **24**, 1679–709.
Wilson, E.O. (1975), *Sociobiology: The New Synthesis*, Cambridge, MA: Belknap Press.
Witt, U. (2004), 'On the proper interpretation of "evolution" in economics and its implications for production theory', *Journal of Economic Methodology*, **11**, 125–46.

3 The place of path dependence in an evolutionary perspective on the economic landscape[1]

Ron Martin and Peter Sunley

1. Introduction: creating space for history in economic geography – the new focus on path dependence

Since the notion of path dependence entered the economics lexicon in the 1980s and early 1990s, particularly through the work of Paul David on the economic history of technology (David, 1985, 1986, 1988, 1992, 1993a, 1993b, 1994), and that of Brian Arthur on nonlinear, self-reinforcing economic processes (Arthur, 1988, 1989, 1994a, 1994b, 1994c, 1994d), it has assumed prominence as an evolutionary concept not only in economics itself, but also across a wide range of social, organisational, technological and managerial sciences.[2] Some even see the concept of path dependence as a major building block of a new interpretative or epistemological paradigm.

Economic geography has also been swept up in this wave of 'path dependence thinking'. A number of leading theorists in the subject have argued that path dependence is one of the fundamental features of the economic landscape. For example, according to Richard Walker:

> One of the most exciting ideas in contemporary economic geography is that industrial history is literally embodied in the present. That is, choices made in the past – technologies embodied in machinery and product design, firm assets gained as patents or specific competencies, or labour skills acquired through learning – influence subsequent choices of method, designs, and practices. This is usually called 'path dependence' . . . It does not mean a rigid sequence determined by technology and the past, but a road map in which an established direction leads more easily one way than another – and wholesale reversals are difficult. This logic applies to industrial locations as well. (2000, p. 126)[3]

Allen Scott (2006) is even more emphatic, and argues that any attempt to understand the economic landscape:

> must formulate the problem by reference to a dynamic of cumulative causation whose logic is definable not in terms of some *primum mobile* or first cause, but in terms of its own historical momentum. This . . . points . . . to the importance of an ontology of regional growth and development that is rooted in the idea of path dependent economic evolution and recursive interaction. (p. 85)

Alongside these and other theoretical invocations, path dependence ideas and phraseology have found their way into a growing body of empirical work in economic geography. Thus the idea has been deployed in discussions of the persistence of regional disparities in economic development; of the 'lock-in' of regions to particular economic specialisations; the revival and reinvention of former local industrial configurations; and of the emergence and self-reinforcing growth of 'high-tech' clusters (see, for example, Bathelt and

Boggs, 2003; Bode, 2001; Boschma, 2004, 2005; Cooke and Morgan, 1998; Fuchs and Shapira, 2005; Gertler, 2005; Grabher, 1993; Hassink, 2005a and 2005b; Kenney and von Burg, 2001; Storper, 1995, 1997). Such economic-geographic studies employing the concept of path dependence reflect a growing interest in the historical dynamics of economic landscapes, a realisation that to understand geographically uneven development, in all of its manifestations, it is necessary to create a space for history.

In one sense, of course, this recognition of the importance of history by economic geographers is not entirely new. The notion of 'cumulative causation', which is closely related to path dependence ideas, enjoyed some degree of popularity within the discipline in the 1970s, though unfortunately it has since largely slipped from visibility. Much Marxist economic geography in the 1980s was concerned to explain uneven regional development as an historical process. For example, Massey's (1984) important work on the spatial divisions of labour was founded on the argument that the economic landscape inherits the legacies of its past development and that these legacies exert an influence on its present and future development. And David Harvey's central aim was (and still is) to explain uneven regional development as a historical process driven by capitalism's episodic phases of accumulation and crisis, as a dialectic between preserving the values of past commitments made at a particular place and time, and devaluing them to open up fresh room for accumulation at some future point in time (Harvey, 1982, 1985, 2006; see also Smith, 1984). But the recent 'evolutionary turn' in economic geography is distinctive in that it draws its inspiration explicitly from evolutionary ideas and concepts, from evolutionary economics, universal Darwinism and even complexity theory, rather than from the meta-narrative of Marxist political economy (see, for example, Boschma and Frenken, 2006; Essletzbichler and Rigby, 2007; Martin and Sunley, 2007). It is within this embryonic subfield of evolutionary economic geography that the concept of path dependence has been accorded particular theoretical and empirical significance. Of especial interest is the question of whether and to what extent the evolution of the economic landscape is a path-dependent process, whether the mechanisms that make for path dependence have a quintessentially local dimension in their form and operation, and thus whether, in this sense, path dependence can be seen as a process or effect that is locally contingent and locally emergent, and hence to a large extent 'place dependent' (Martin and Sunley, 2006).

However, as we argued in our previous conspectus of the concept (Martin and Sunley, 2006), despite the increasing use of path dependence terminology and notions by economic geographers, there has been little extensive or detailed discussion of what the ontology called for by Scott might look like. In fact, economic geographers have tended to apply the concept of path dependence as if it is self-evident and wholly unproblematic. As Glasmeier (2000, pp. 269–70) complained, in economic geography the concept of:

> path dependence is often invoked uncritically as an explanation for a particular industrial [and, we might add, regional] experience. Usually lying behind the notion of path dependence is a series of factors that together add up to a directional bias. Just exactly what provokes path dependency is rarely communicated, however; this often erroneously leads to uni-dimensional invocations of the term.

We would go further and suggest that the very idea of 'path-dependent economic evolution' as advocated by Scott is itself in need of careful interpretation and conceptualisation if it is to serve as the basis of a meaningful ontology of regional growth and development. The stakes here may be high. Path dependence may help explain why regional growth disparities persist; it may help explain why particular industries and technologies develop in certain locations but not in others; and it may help us to understand why some regional economies are better able to adapt over time than others. Our previous paper (Martin and Sunley, 2006) was intended as an initial exploration of such issues. In that paper, we explored the sorts of processes that could give rise to path-dependent economic development in a geographical – and specifically, regional – setting, and sought to argue that many of those processes are themselves place-dependent. The aim here is to build on that discussion and to focus much more specifically on how far and in what ways path dependence can serve as an *evolutionary* concept for studying the economic landscape. The thrust of our argument is that this task is problematic. For one thing, we find that despite its declared emphasis on the importance of history, path dependence theory as formulated by its leading architects – David, Arthur, and others – retains elements of equilibrium thinking, which we contend is in tension with the idea of path dependence as an evolutionary concept. We are sceptical that the evolution of the economic landscape ever tends towards any equilibrium state: to be sure, the economic landscape is characterised by self-organisation and order, but these are not the same thing as equilibrium. We argue, therefore, for a wider view of path dependence that allows for patterns and trajectories of development that do not approach or reach an equilibrium state, and that do not require an equilibrium interpretation. This leads on to a discussion of how to characterise such paths and their emergence in a regional context. Our conclusion is that a rethinking of the notion of path dependence may be called for if it is to function as a core concept within an evolutionary perspective on the economic landscape. To this end, economic geographers could benefit from looking at the wider discussions of path dependence and historical change that have taken place in recent years across the social, political and policy sciences, and at related ideas in evolutionary ecology.

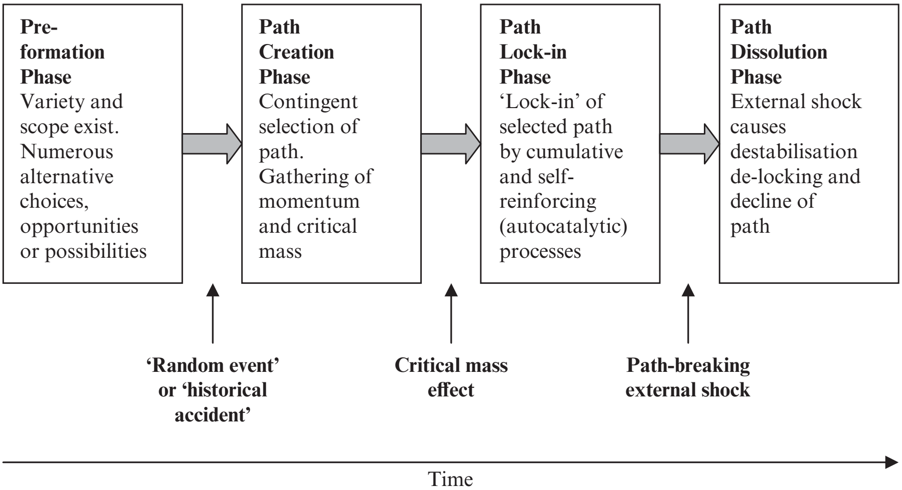
2. The basic model of path dependence and its geographical application

One of the abiding intentions behind Paul David's development and proselytisation of the concept of path dependence has been to persuade economists – and especially neoclassical economists – to move beyond traditional equilibrium modelling and to 'take history seriously'. This endeavour resonates strongly with evolutionary economists, some of whom, such as Hall (1994), have elevated the notion of path dependence to the status of a 'first principle' of evolutionary economics. David (2001, 2005, 2007) himself argues that path dependence is a property of a 'wide array of processes that can properly be described as evolutionary', including economic processes. But in what sense, precisely, does path dependence function as an evolutionary concept? Neither David nor Arthur, the primary exponents of path dependence ideas, has set out a fully specified model of economic evolution based on path dependence, nor discussed how the concept relates to other key ideas in evolutionary economics. Nevertheless, it is possible to infer from their various writings the sort of model of economic evolution implied by their depiction of path dependence.

According to Sydow et al. (2005), the basic model of path dependence developed by David and Arthur implies a three-stage model of the historical (and, we might add, geographical) development of a technology, industry, institution or organisation. In the first, 'pre-formation' phase, as they term it, considerable scope and variety exist for exploring or developing a new technology, product, industry, or institution (and again, we may add, a location for such an activity). In David's and Arthur's work the search process is portrayed as largely undirected, and decisions mainly contingent. There may be several alternative opportunities being explored at this stage. A 'critical juncture' then occurs – in David's accounts this is usually a 'historical accident' or 'random event' – that results in one particular development opportunity or decision being 'selected' or preferred over alternatives, a process that David likens to the idea of 'bifurcation' used in complexity theory. This opportunity or development then begins to attract other actors, or acquires market influence, and a critical mass around this activity begins to build up and a development path is formed. Once this critical mass achieves a certain size or momentum, the path gets 'locked-in', and a third phase of cumulative and self-reinforcing development along this path ensues.

Sydow et al. criticise the basic David path-dependent model on several grounds. For example, they argue that the assumption that path creation is a random event or 'accident of history' ignores the fact that the emergence of a new technology, a new product, or a new organisation, is often the outcome of purposive behaviour and directed decision-making by economic agents (see also Garud and Karnøe, 2001; Martin and Sunley, 2006). Further, they suggest the model is incomplete in the sense that it says nothing about how paths 'de-lock', break up and dissolve. These are obviously important aspects of the path dependence idea, and clearly crucial to its relevance as an evolutionary concept. While it is true that David tends to see the emergence of a path in terms of some 'random event' or 'historical accident', it is not strictly true that he says nothing about how a path ends. In fact he has a very specific view on this, namely that the break-up and dissolution of a technological, industrial or institutional trajectory is brought about by an 'external shock' of some sort, which then destabilises the system and opens up opportunities for a new path to emerge. Thus, the basic path dependence model actually posits four stages of the development of a technological, industrial or institutional trajectory: pre-formation, path creation, path lock-in, and path dissolution (see Figure 3.1).

There are two main ways in which economic geographers have used a basic path dependence model of this sort: to explain the evolution of a particular industry, technology or institution either in a *given* location (region, city), or *across* locations. In the former case, interest has focused on identifying the initial locally 'contingent' factors or stimuli responsible for the emergence of the industry, technology or institution (the creation of the path) in the area in question, and on the types of self-reinforcing mechanism and co-evolutionary processes that explain its subsequent path-dependent development. In particular, attention is often directed to the role of local 'network externalities' (to use David's terminology) or 'increasing returns' (to use Arthur's). In some applications, particularly in relation to high-tech regions, the local embedding of such network externalities and increasing returns is viewed as key to innovative and competitive success. In such cases, 'lock-in' is seen as a positive process. In other applications, however, typically older industrial regions, 'lock-in' is seen as a negative feature, whereby a region or locality becomes over-reliant on, or dominated by, a particular self-reinforcing industrial-technological path that renders the regional or local economy increasingly structurally and technologically rigid, restricting thereby its capacity to adapt to changing competitive forces ('the weakness of strong ties' argument). In this version the basic model is often used to argue that the path-dependent overspecialisation of regions is likely to make them



Source: Adapted from Sydow et al. (2005).

Figure 3.1 Phases of economic evolution of an industry or technology implied by basic David–Arthur-type path dependence models

vulnerable to sector-specific (and hence region-specific) shocks that then bring about the decline of the industry in question, and with it much of the regional economy as a whole.

In the second type of application of path dependence in economic geographic work, the focus is less on the path-dependent evolution of a particular industry in a particular region, than on how a given industry evolves spatially across a multi-region or multi-location economic landscape (although of course the industry may end up entirely concentrated in just one particular location). In a similar way that in the basic path dependence model a new technological or industrial path is assumed to be initiated by random, chance or happenstance events, so in this geographical version of the model the parallel assumption is that initially there is spatial indeterminacy as to where a new technology or industry will begin to emerge. Several possible potential locations may be equally suitable initially. In which of these locations the industry or technology is triggered will be largely an accident of history – a situation described by Boschma and van der Knaap (1997) as one in which the 'windows of local opportunity' are 'open'. Which locations then get 'selected' for subsequent path dependence development will depend on small arbitrary differences in the conditions and circumstances in those 'trigger' locations, and whether and to what extent cumulative self-reinforcing feedbacks emerge:

> the discontinuous nature of major innovations . . . implies that the spatial formation of new industries involves spontaneity or indeterminacy because it is unlikely to be determined by or bound to particular places . . . the actual outcome depends on small arbitrary events, magnified by a positive feedback mechanism, which, in our approach, is achieved by the creative ability of firms that build up a favourable local production milieu around them. (Boschma and van der Knaap, 1997, p. 182 and p. 187)

The parallel with Arthur's path dependence model of industrial location is clear:

> [E]arly firms are put down by historical accident in one or two locations; others are attracted by their presence, and others in turn by *their* presence. The industry ends up clustered in the early-chosen places. But this spatial ordering is not unique; a different set of early events could have steered the locational pattern into a different outcome . . . We might call this view *historical dependence*. (Arthur, 1994a, p. 50)

Thus the eventual spatial pattern of an industry is interpreted as being the outcome of early (chance) events and subsequent spatially selective path-dependent cumulative processes. Models of this sort have tended to focus overwhelmingly on how new industries emerge and develop across space, and have had little to say about how a given spatial-industrial path dissolves, though the implication of the windows of locational opportunity concept is that each new technological innovation opens up new windows and hence new spatial configurations of economic development, so that 'the long term evolution of the spatial [economic] system is potentially unstable' (Boschma and van der Knaap, 1997, p. 198). How this relates to the observation that long-run patterns of relative regional prosperity are often highly persistent (and path-dependent) over long periods of time is not clear.

While these uses of the basic path dependence model in economic-geographic work are certainly suggestive, they leave several issues still far from fully resolved (see, for example, Table 3.1). Many of these have to do with how we conceptualise the meaning and nature of path dependence within regional and local settings. Others relate to the sources of path dependence and how far these are shaped by local conditions and circumstances, that is how far path dependence is itself place-dependent. Still others concern why the degree of path dependence seems to vary across the economic landscape. And, importantly, there are basic questions about what sort of economic evolution is implied by path dependence: does it imply slow, incremental change and development, or a more punctuated pattern, of successive periods of relative stability, even inertia, of economic structures and technological development interrupted by episodic bouts of radical industrial and technological change? It is on these latter issues that we wish to focus in what follows.[4]

3. In what sense an evolutionary model of the economic landscape? Path dependence versus equilibrium

In much of economics, equilibrium rather than history has been the central organising concept in theoretical and empirical enquiry. The quintessential feature of this 'equilibrist methodology' (as Setterfield, 1997, calls it) is that the development of the economy is interpreted not as being shaped in any significant and persisting way by particular events that occurred in the historical past, but as a movement towards a hypothetical equilibrium outcome. The aim is to demonstrate how, under specific given (and typically highly abstract, simplified and idealised) assumptions (as to consumer tastes, technological knowledge, the nature of competition, institutional arrangements, etc.), an economy tends ineluctably and deterministically towards a limiting unique ex ante equilibrium state that is invariant over time and space. The economy, in other words, is conceptualised as an *equilibrium process*, which (following Harris, 2004) might be depicted as:

$$\textit{Equilibrium process: } x(t+1) = F_x^{\,e}\,(x(t)),\ -\infty \leq t \leq +\infty;\ x^e = \text{equilibrium point}$$

Table 3.1 Some key issues in the application of the concept of path dependence in economic geography

Key questions	Issues
What is the object of study?	What precisely are we referring to: a region's firms, its industries, or the regional economy as an aggregate? What is the relationship between the micro-level and the regional aggregate level? Path dependence is a multi-scalar process, operating at different levels. How do these interact within regional economic spaces? Is regional path dependence an emergent effect?
What are the mechanisms of regional path dependence?	What are the sources of path dependence within local and regional economies (sunk costs and infrastructures, external economies of industrial localisation, general agglomeration economies, local socio-cultural-institutional embeddedness, local and extra-local economic dependencies and linkages)? How do these mechanisms vary (and interact) across space?
Do all parts of a regional economy display path dependence?	Is a certain threshold of interaction and inter-relatedness within a region economy required before it displays aggregate path dependence? Can different paths co-exist within a region, and how do these interact – what is the nature of 'inter-path coupling'?
What are the sources of regional path dependence?	Some will be more or less specific to particular types of regional economy; but different forms of path dependence are likely to co-exist within a regional economy. In this sense, the notion of regional path dependence is complex and multi-dimensional. Is a single, overarching theory possible?
Do the nature and strength of path dependence vary from region to region?	Regions differ markedly in economic structure, institutions, connections to other regions and beyond, etc., so the nature and degree of path dependence might be expected to vary from region to region. In what ways is path dependence a locally contingent – that is, a place-dependent – process?
In what sense does path dependence explain the evolution of the economic landscape?	What sort of evolution is implied by path dependence? Is path dependence consistent with incremental economic change and mutation? Or does it imply a 'punctuated equilibrium' process of evolution and transformation of the economic landscape? Where do new paths come from, and why do they emerge where they do?
How pervasive is path dependence in the economic landscape?	Is it a fundamental feature of regional development and evolution – i.e. inevitable and indeed necessary for regional growth to take place? Or is it more typical of economically specialised regions and localities, and less likely to emerge in areas with diverse economic-technological structures?

Source: Based on Martin and Sunley (2006).

where x is the economic state or outcome of interest, and the function F_x^e, governing the change over time of the system, generates a unique and stable equilibrium x^e. In such a conception, equilibrium is not a real emergent outcome of actual historical processes and events, but an abstract solution state determined by the specific assumptions, equations, and exogenous parameters and variables built into a formal economic model, the basic purpose of which is to determine the existence and stability of equilibrium.[5] Any notion of 'dynamics' in such models concerns only the model economy's movement towards its equilibrium: once in that equilibrium state, the system is in stasis. Equilibrist economics is basically antithetical to notions of history and evolution (Harris, 2004; Kaldor, 1934, 1972, 1985; Robinson, 1974; Setterfield, 1995, 1997).

In contrast to equilibrium economic processes, in an economy characterised by path dependence the specific details and sequence of historical events govern the unfolding course of development – what David calls 'historically contingent evolution'. Following Harris (2004) and Page (2006), such a process might be expressed as:

$$\textit{Historical process: } x(t+1) = F_x(t)(h(t)x),\ -\infty \leq t \leq +\infty,\ h(t)x = x(t), x(t-1)\ldots x(0)$$

where $h(t)x$ is the history of past outcomes of x from $t = 0$ up to time t, and the function $F_x(t)$ maps that history into the next outcome.[6] The outcome function can itself change over time so it is indexed by t. In such a system, the present state of the economy will depend on where it has come from, and on how it got there: this is what is generally meant by path dependence. There is nothing inherent in such an historical process that necessitates that it possesses or reaches a stable equilibrium state. Indeed, the concept of path dependence can be argued to be fundamentally antagonistic to an equilibrist methodology. According to the latter, the long-run equilibrium state can be defined and reached independently of the path taken towards it, whereas with path dependence any long-run configuration that is reached by the economy will *depend* on the path taken towards it. There is thus no predetermined economic outcome, no outcome independent of history or context. To argue that an economy, an economic landscape, is a path-dependent historical process would thus seem to be incompatible with simultaneously arguing that it is an equilibrium process.

Yet, while emphasising the need to move beyond the equilibrist methodology of mainstream economics, Paul David and other path dependence theorists nevertheless seem reluctant to relinquish the idea of 'equilibrium thinking' altogether.[7] In fact, recently, David has explicitly referred to his approach to historical economics as 'path-dependent equilibrium analysis' (David, 2005a, p. 153). His strategy to reconcile the apparent contradiction in this phrase is to define path dependence in terms of the dynamics associated with particular types of non-ergodic stochastic processes and systems that possess a multiplicity of limiting distributions, that is multiple equilibria:[8]

> The elaboration of theories around the core concept of path dependent dynamics . . . encourages and enables economists to entertain the possibility that, in place of a unique equilibrium-seeking dynamic, they should envisage a process that is seeking an evolving and historically contingent equilibrium. (David, 2005b, p. 2).

Under such conditions,

> Small events of a random character – especially those occurring early on the path – are likely to figure significantly in 'selecting' one or other among the set of stable equilibria, or 'attractors' (David, 2007, p. 151).

Which of these multiple equilibria is reached or 'selected', it is contended, will depend on the initial state of the system – on initial 'random events' – and the chains of transitions produced by repeated iteration of the system over time. In this sense, there is no single ex ante unique distribution of the economic system, as in general equilibrium economics: the actual limiting distribution depends on history, and particularly where the economic system (for example, the technological or industrial structure) started. Further, according to David, once the processes of path dependence have locked the system (technology or industry) into one of these alternative stable equilibria (or 'attractors'), it requires an external shock to break the particular development path or trajectory in question. The evolution of the technological, industrial and institutional structure of an economy in this path dependence model would thus seem to be akin to 'punctuated equilibrium', a (stochastic) pattern of historically contingent evolution in which particular paths of technological, industrial and institutional development become 'selected', locked-in to stable equilibria, and, at some point, eventually disrupted and broken up by external shocks. As David puts it:

> Sudden shifts in structure, corresponding to the new evolutionary biologists' notion of 'punctuated equilibria' . . . may open up a way for the formulation of dynamic models that are compatible with 'stage theories' of development. (2005a, p. 187)

What is interesting about this 'path-dependent equilibrium analysis' is that it resonates with how notions of 'path dependence' and 'history' have been used in the so-called 'new economic geography' (NEG) that has arisen over the past two decades. Adherents of this approach to analysing the space economy are explicitly equilibrist in orientation, asserting that it is 'the general equilibrium modelling of an entire space economy which sets the approach apart from traditional location theory and economic geography' (Fujita and Mori, 2005 p. 380). Moreover, while committed to this equilibrist representation of the economic landscape, NEG theorists simultaneously claim to recognise that 'history matters' and that their models incorporate 'path dependence'. The way this is achieved is precisely by constructing (deterministic) models that yield multiple equilibria outcomes (in this case equilibrium spatial patterns of industry and employment) depending on the 'initial conditions' (such as the level of transport costs, the relative mobility of labour and capital, the initial distribution of industry between regions, and the number of regions) specified in the models.[9] For example, the basic two-region 'core–periphery model', the foundation of NEG theory, possesses five such possible equilibria (three stable and two unstable). Because the 'initial conditions' determine which of the alternative equilibrium landscapes emerges, NEG theorists feel able to claim that 'history matters' in their models. And because these models generate their equilibrium outcomes via processes of self-reinforcing agglomeration processes, it is also claimed that they embody path dependence (or 'locational hysteresis').

Now the notion that 'initial conditions' matter is certainly an improvement over the timelessness of conventional equilibrium economics, since potentially it at least draws attention to one aspect of the past history of a system in the determination of future outcomes. However, to our mind, characterising path dependence as a problem of multiple equilibria (whether stochastic, or deterministic as in NEG models) is to restrict the concept and its relevance for analysing the evolution of regional and local economies. In David's work, as in NEG models, multiple equilibria are conceived as representing a system that possesses a variety of 'locally stable attractors' (to use his term), from which the final equilibrium position is 'selected' on the basis of the system's initial starting position or state. In effect, if we know the various alternative possible starting states, the array of possible final outcomes to which they may conform becomes determined a priori, and all that remains to be ascertained is which outcome of the possibles will actually be selected. To quote Setterfield:

> Characterising path dependence as a problem of multiple equilibria constitutes a limited conception of this phenomenon because it overlooks the possibility of a path dependent process *creating* its own set of final outcomes in the course of its evolution. The basic distinction here is one of ontology. In a selection process, the environment comprises a pre-prescribed or given external reality to which decision makers must adapt (through a process of selection). In a creation process, however, the environment is not pre-prescribed, but is instead 'open' – it remains to be constructed and defined in the course of the concrete functioning of the system. (Setterfield 1997, p. 64, emphasis as in original)[10]

What is being argued here is that even the idea of a path-dependent equilibrium would seem to run counter to the basic principles associated with historical time. However an equilibrium is defined, it is a state from which a system will display no endogenously generated tendency to deviate. External shocks, as in David's schema, may disrupt and destroy that equilibrium, but these are extraneous influences. What this suggests, then, is that once we are in equilibrium, history and change effectively end: the future is predetermined by state of the system corresponding to the equilibrium that has been achieved. Indeed, a commonly held view of path dependence is that it implies, or refers to the situation where, the system gets locked into an inertial state. To define path dependence as the historically contingent selection process between multiple equilibria seems to imply that path dependence only matters in the selection and movement towards an equilibrium position, and that once in that equilibrium position path dependence merely serves to reproduce that equilibrium, that is to entrench stasis, a position of no change.[11]

Setterfield's discussion of the relationship between path dependence and equilibrium is one of the most detailed we have been able to find. However, although he argues forcefully for the incompatibility of history and equilibrium, even Setterfield cannot bring himself to dispense with equilibrium thinking. He suggests – unsuccessfully in our view – that the notion of equilibrium can still be usefully retained in a historical or path dependence approach to economic theorising under two possible strategies (Setterfield, 1997). The first is to treat equilibrium in the economy as a 'pedagogic device' to 'lock-up without ignoring' the complexities of historical time. But surely this is to introduce a form of conditional closure to what in reality is an open, historical system, and seems to us to be little different from the 'ceteris paribus' position adopted by mainstream equilibrium economics. In effect, it is to assume that the various mechanisms and forces that produce change in the economy are 'held constant'. As Setterfield himself states:

> The pedagogic use of equilibrium, designed to 'lock up *without ignoring*' pertinent features of historical time, suggests that in some circumstances it may be strategically useful to demonstrate what it would mean for the economy to be in 'equilibrium', even if historical motion and change are believed to be pervasive. (Setterfield, 1997, p. 68)

This is difficult to reconcile with an interest in economic evolution if indeed that is the focus of our interest. His second suggestion is to treat equilibrium as the actual 'temporary' outcome of a path-dependent process that may yet give rise to a subsequent endogenous process of 'innovating out' of equilibrium. Path dependence then is defined in terms of a sequence of 'temporary' equilibria. Setterfield contends that if all equilibria are conceived as essentially temporary equilibria, then 'the antagonism between historical time and the concept of a state of equilibrium is at least attenuated' (1997, p. 68). A not unrelated take on this issue is to be found in Bassanini and Dosi (2001) in their discussion of path dependence. Having given the idea of asymptotic path dependence a multi-equilibrium interpretation, they go on to recognise the need to 'broaden the definition of path dependence to encompass the case where there is no convergence [lock-in] to any asymptotic behaviour' (p. 54). However, like Setterfield they view this alternative form of 'path dependence without asymptotics' as a sequence of 'temporary' resting states or equilibria:

> Actually, many examples, from the evolution of institutions, organisations and technologies, suggest a world wherein temporary resting states are 'metastable' in the sense that, on a longer time scale, that are persistently overcome by new development leading to new 'temporary resting states'. (Bassanini and Dosi, 2001, p. 54)

While Setterfield's idea of 'endogenously innovating out of equilibrium' is a useful corrective to the usual appeal to the necessity of an exogenous shock to move the economic system out of an equilibrium state, we still fail to see why the notion of 'temporary' equilibria or resting states is needed. How do we (or economic agents themselves) know an economic system is in a state of 'temporary' equilibrium? And how short in duration can a position of 'temporary' equilibrium be before it becomes indistinguishable from a process of incremental endogenous change? Or, on the other hand, how long in duration before it becomes indistinguishable from the idea of punctuated equilibrium.[12]

In our view, the very idea of equilibrium – whether path-dependent or not – is difficult to reconcile with what is generally taken to be the defining feature of an economy viewed as an evolving system. According to Witt (2003, 2006), for example, the key focus of evolutionary economics is on the processes and mechanisms by which the economy *self-transforms itself from within*, that is on processes of endogenously generated change.[13] This conception is fundamental to what Castellacci (2006) calls an *evolutionary ontology*:

> The co-existence of random and systematic factors driving economic evolution . . . and the combination of inertial and dynamic forces, both constitute important elements in an attempt to explain the most important stylised facts about economic evolution . . . the existence of structural change (the old Schumpeterian 'creative destruction'), persistent differences in growth rates between regions and countries, phenomena of path dependency, and cumulative causation patterns. Such real phenomena are regarded as unique events in historical time. Differently from the neoclassical metaphor of a steady state, evolutionary scholars describe an ever-changing and never-ending process of growth and transformation. (p. 869)

Evolutionary thinking requires a shift in mindset from the characterisation of equilibria to the specification of dynamic and historical processes. Economic evolution is a species of 'far-from-equilibrium' process and what keeps economies far from equilibrium is the particular set of knowledge-generating and application processes that define the modern economy. Since knowledge is never in equilibrium, it is difficult to envisage an economy ever being in equilibrium (Ramlogan and Metcalfe, 2006), and its behaviour cannot be understood by asking whether it has any equilibrium or steady states (Durlauf, 1997).[14] Both David and Arthur (and several other scholars of path dependence) recognise the role of learning as a causal mechanism of path dependence, but at the same time operate with an equilibrium interpretation of path dependence. Further, complexity theory tells us that 'far-from-equilibrium' systems can be highly structured and patterned, and strongly path-dependent.[15] In such complex systems, structures and patterns emerge from processes of self-organisation not as outcomes of asymptotically equilibrating mechanisms.[16]

It would seem, then, that we can distinguish at least three conceptions of path dependence as an evolutionary concept. The first and most restrictive is the David-type conception that defines path dependence in terms of the historically contingent selection of, and lock-in to, one of a multiplicity of possible stable equilibrium outcomes. Once locked in, it requires an external shock of some kind to dislodge or de-lock the equilibrium state. The process of economic evolution implied is one in which a technology or industry – or regional economy – becomes locked in to a particular selected equilibrium path or state, in which state it remains until such time that an external shock or disturbance (such as a radical new technology, or a major upheaval in the market for a region's product(s), etc.) disrupts that state (see the highly schematic representation in Figure 3.2).

It is perhaps significant that all of the primary examples of path dependence cited by David (the QWERTY keyboard, VCR video, AC electrical current, light-water reactors) have been of technologies or technological configurations that once 'locked in' remained largely unchanged. These might be interpreted as examples of 'stable equilibrium' states, but they are clearly only one type of economic evolution, and arguably a restricted form at that.

The second conception is of the Setterfield type, wherein path dependence processes generate a 'temporary equilibrium' outcome that then gives rise to a subsequent endogenous process of 'innovating out' of equilibrium. According to Setterfield, the very fact of the economic system in question (an industry, say) being in a state of 'temporary

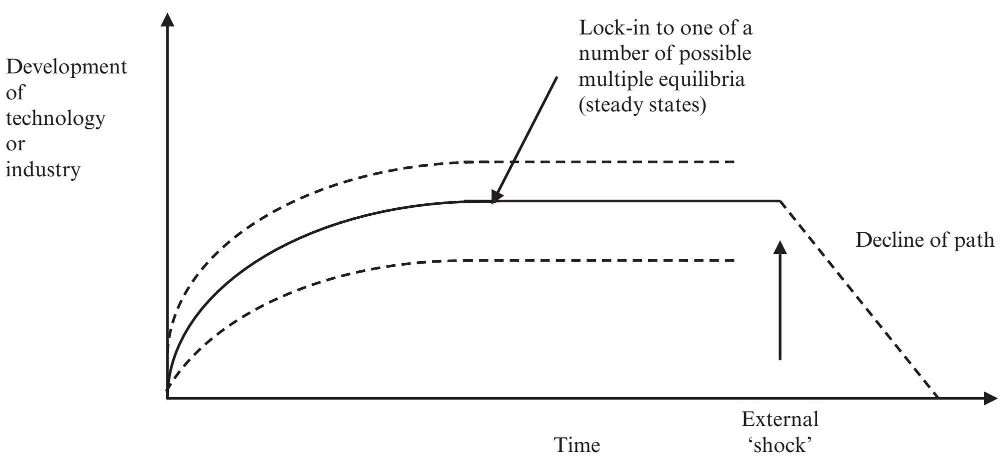


Figure 3.2 The David-type model of path-dependent evolution

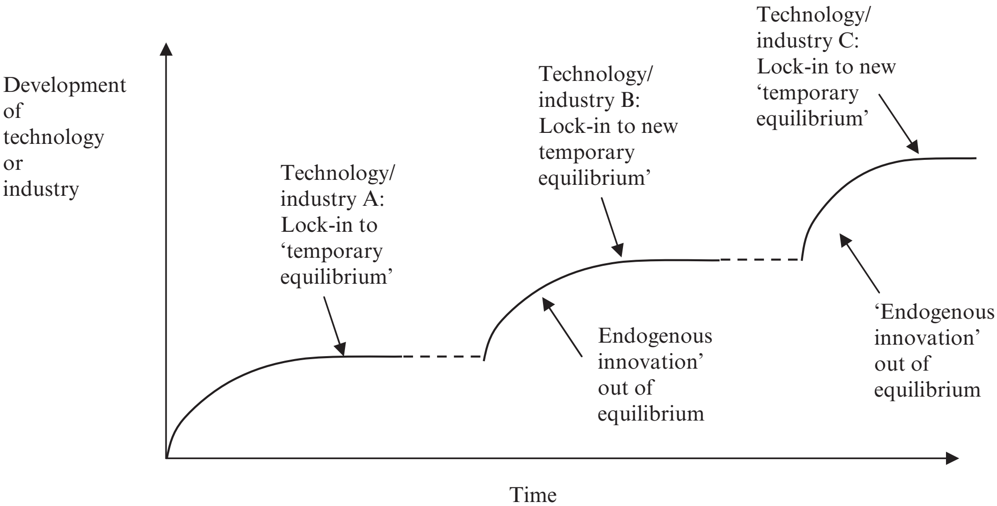


Figure 3.3 A Setterfield-type model of path-dependent evolution

equilibrium' itself will tend to stimulate purposive behaviour by some economic actors to explore pre-emptive breaks from the locked-in technological-industrial activity in order to establish a new competitive 'temporary equilibrium' path of technological or industrial development. Path-dependent economic evolution in this schema is one of a succession of 'temporary' equilibria' (Figure 3.3). We are not overly convinced by this conception, since any temporary period of stability would be interpreted necessarily as an 'equilibrium', when the system may in fact be far from any such state.

A third conception is that which frees the idea of path dependence from any necessary connection with equilibrium, and which views path dependence as a dynamic open historical process by which technologies, industries and institutions evolve along unfolding trajectories: what we might term a non-equilibrium conception of path dependence (Figure 3.4).[17] These trajectories are shaped not only by the sequences of prior developments and influence of earlier events, but also by the evolution of the processes (mechanisms) of path dependence themselves. This is what the time-varying historical process function $F_x(t)$ used above is meant to capture. Under this approach, explicit allowance is made for the possible interaction between the evolving technology, industry or institution and the various path dependence processes (such as learning, network externalities, etc.) that are shaping it. Put another way, a path-dependent technological, industrial or institutional trajectory and its associated path dependence processes may co-evolve. Such a system need never approach any form of equilibrium. Furthermore, such a schema allows for various possible evolutionary pathways. It can encompass the case of endogenously generated incremental path-dependent evolution as well as continuous evolution in response to constant change in the external environment: it allows for both slow and more rapid path-dependent evolution. It allows for the adaptation and mutation of a technology, industry or institution over time (of which numerous examples abound – indeed many technologies, industries and products are characterised by this mode of evolution). And it also includes those cases where a path-dependent

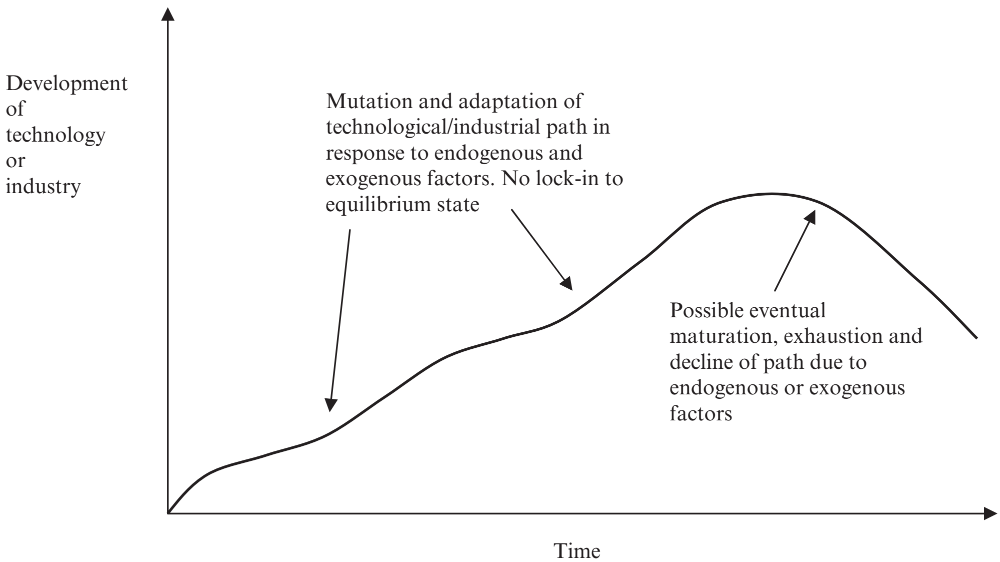


Figure 3.4 An open, non-equilibrium-type model of path-dependent evolution

form of development endogenously generates the mechanisms of its own relative decline and dissolution.

These different possibilities seem to us to be highly relevant to discussions of the path-dependent evolution of local and regional economies. But releasing the notion of path dependence from any necessary connection with equilibrium states in turn raises a wide set of issues about exactly how we understand evolution along an industrial or technological path.

4. Exploring non-equilibrium path dependence: economic evolution along paths

As we have seen, the notion of path-dependent equilibrium is based on the view that the operation of feedback processes gradually increases the probability of a certain (equilibrium) outcome. These processes cumulatively reinforce the direction of change. In contrast, however, an open non-equilibrium interpretation of path dependence needs to recognise that while institutions and past trajectories channel change, at the very same time they also enable adaptation and the launch of new paths. Indeed, in some cases the very same mechanisms alleged to produce path-dependent equilibria – increasing returns, learning, network externalities and the adoption and spread of technological standards – are precisely those that can act to maintain and even increase the probability of adaptation and departures to new paths. We would contend then that a realistic theory of path dependence needs to examine the conjunction of both sets of constraining and enabling effects.

This non-equilibrium perspective has important implications for our understanding of industry and product life-cycle approaches that attempt to explain evolution along particular paths. According to product cycle theory, as products become more standardised and routinised, their production location shifts from creative seed-beds and core cities, where product innovation first occurs, towards lower labour cost locations in more peripheral regions and countries (see, for example, Vernon, 1966). The predictable evolution along a path determined by the maturity of the product thus leads to the dispersal of production and growing wealth in lower cost regions. This evolution towards predictable types of development path is, of course, highly problematic (Markusen, 1985; Storper, 1985; Taylor, 1986). In Storper's (1985) evaluation, for example, the product cycle approach was too essentialist as it sought regular mechanical, and closed system relationships among variables, and failed to recognise the full open-ended ontology of economic processes. In particular, he suggested that industrial 'life-cycle' notions extrapolate from transitory and contingent empirical phenomena and credit them with the status of a teleological developmental logic. The choices in manufacturing and business strategy envisaged are too determinate and too closely related to the extent and maturity of the market: 'the cycle conceives of technological history as a series of determinate, repeating events, which it assumes to be a common to all industries. Thus it would turn economic history into a kind of "natural history"' (Storper, 1985, p. 271). Storper advocated a more open-ended view of history as a series of unfolding events and in this way he dismissed evolution as an endogenous organic unfolding and instead gestured towards a more neo-Darwinian view of industrial change. However, while his argument leant towards a more idiographic and unpredictable understanding of industrial and regional paths, a completely open-ended approach risks losing sight of path dependence altogether.[18]

More recent explanations of industrial life-cycles certainly identify generic forms of path dependence in a range of industries. There are several different varieties of the industry life-cycle approach but they share a similar schematic interpretation of how industries develop (Agarwal, 1998; Agarwal and Gort, 1996; Klepper, 1997). Most accounts distinguish between an early formation stage, an intermediate development stage and a mature stage. During the formative stage, firms enter a new industry by producing a relatively new product and face a great deal of uncertainty. During the second stage, process techniques become more refined, and markets and outputs grow rapidly. During the mature stage, production techniques are further refined, market growth levels off and significant innovations are fewer. In terms of population dynamics, the entry of new firms is greatest during the formation stage. Typically it is argued that a sharp drop, or shakeout, in the number of producers occurs during the development phase and the number of firms continues to decline in the mature stage (Klepper and Simons, 2005). Klepper (2002) argues that innovation also tends to evolve during the course of an industry's evolutionary life-cycle. Innovative activity is greatest during the early formation stage and new and small entrants have an innovative advantage. In the development phase, there is a reduction in innovative activity and a shift towards established large enterprises. In the mature phase, any new entrants face a comparative technological disadvantage. Audretsch and Feldman (1996) spell out some geographical implications and argue that the propensity for innovative activity to be geographically clustered is strongly related to the stage of the industry life-cycle. Innovative activity tends to be more clustered in the early and formation stages as small new entrants benefit greatly form localised knowledge spillovers.

The stylised facts of industry life-cycles suggest that path dependence arises from several different forces (Klepper, 1996, 2002; Klepper and Simons, 2005). In one approach a radical invention induces firms to enter a new industry. Those that are successful innovators expand their production, and the pressure of output growth on prices causes the less successful innovators to exit the industry. In another approach, firms enter an industry with alternative and competing product designs. During the course of time one design becomes the 'dominant' or standard design. Subsequently, opportunities to enter the market based on new designs are diminished and firms compete on the basis of process innovations. In other words, the lock-in of the dominant design reduces opportunities to enter (Klepper, 1996). In the third approach, described by Klepper (2002), prices tend to fall as output in an industry expands, and this limits the ability of later entrants to catch up with the size of earlier entrants. But larger firms benefit most from R&D as they can apply its benefits to the largest amounts of output. As price continues to fall, therefore, the smallest firms and least able innovators are driven out of the market. While this debate continues, it is clear that all these varieties of explanation rely on forms of path dependence. Whether it is the lock-in of a dominant design through network externalities and learning effects, or the positive feedback from R&D and learning-by-doing advantages to first movers, these dynamics envisage types of cumulative process. They all suggest that the sequencing of firm entry and the accumulation of capabilities through time are crucial to industrial evolution, and hence to the path-dependent development of local and regional economies.

A non-equilibrium interpretation emphasises that such cycles are not deterministic. In the first place, industrial trajectories are not exclusively determined by their product market maturity. For example, Hudson (2005) argues that the trajectories of industries in north east England were 'path contingent' as they depended not only on trends in international markets but also national political decisions. Second, learning by experience and the accumulation of resources actually facilitate actors' reflexive adaptation (Rantisi, 2002). As a consequence, important departures from the life-cycle model have been found in some industries as new markets and niches are repeatedly constructed and exploited. For example, Klepper and Thompson (2006) suggest that the continual rise in the number of entrants in the laser industry may be because of the importance of specialised submarkets. The proliferation of submarkets provides new niches for entrants and, as the nature of the technology does not allow significant economies of scope, established incumbent firms cannot dominate these new markets. Where new entrants can occupy distinctive market niches and differentiate themselves from incumbent firms, they prove resilient and sustainable. Further evidence of enabling forms of path dependence is that product innovation does not always wane in the mature phase of certain industries. Instead, it may rise sharply. For example, in the mature automobile industry, Japanese and some European producers gained an innovative lead over American producers through both process and product innovations (Heffernan, 2003; Klepper, 1997). Such cases of 'resurgent innovation' have been widely described as evidence of a phase of 'dematurity' (Storper, 1985). Processes of this sort would seem to fit a path-dependent model in which an industry (and that part of a local or regional economy dependent on it) undergoes radical adaptation, in effect renewing (or extending) the industry's development path.

In order to explain why some mature sectors succumb to negative lock-in and complacency while others demonstrate increased innovativeness, it is crucial that we examine the intersection of enabling and constraining processes. For instance, cognitive path dependence means that cognitive priorities and frames are typically less than fully rational and tend to be inductively based on past experience (Beinhocker, 2006). There is some evidence that most firms exhibit path-dependent learning (Chapman, 2005; Glasmeier, 2007), and in many cases more radical innovations are pursued and developed by smaller start-up firms. However, in some cases, new firms may also stimulate incumbent firms to diversify and adopt new technologies (Malerba et al., 1999). Given this, the stylised general paths of industry evolution can be interpreted as the outcomes of a continuous tension between life-cycle processes and those that yield concrete variations and departures from standard life-cycle patterns. Furthermore, another source of non-equilibrium tendencies may be the place-specific interactions between different industrial paths.

In the case of a region or locality dominated by a single industry, then, the product and technological maturity of the industry will obviously play a key role in shaping the path of the relevant geographical economy. Yet, in all probability, such a simple and direct influence is likely to be unusual, for several reasons. First, as several authors have noted, even in the case of a single industry prone to clustering it has been found that different clusters show different paths and life-cycles (Menzel and Fornahl, 2007). For example, Saxenian (1996) showed that, because of variations in local firm strategies and network structures, the Boston Route 128 computing cluster declined just as the Silicon Valley cluster expanded. Second, the translation of industrial maturity into cluster and regional development is likely to be complicated by the heterogeneity of responses of different firms, and by cross-scale interactions between firms, clusters, and urban and regional contexts. Where firm adaptation is short-termist and primarily aimed at cost-cutting it may effectively undermine the adaptability of larger scale entities (Chapman et al., 2004; Sunley, 1992).[19] In such cases, there are local conditions and place dependencies that shape the competitiveness of different clusters so that an understanding the state of the industrial life-cycle is necessary but not sufficient. Moreover, most high-technology clusters are actually composed of several technological trajectories and the switching of resources between them may be especially important in times of crisis (Bathelt, 2001). Most life-cycle approaches, however, do not consider the possibility that the co-location of different industries in a particular region makes a real difference to the co-evolution of their trajectories. It is entirely possible, of course, that the different industries and activities that make up a given region's economy evolve wholly independently of one another. It could even be that each of a region's industries evolves in a path-dependent manner but is uninfluenced by the development paths of the other local industries: a situation we have called 'multiple *unrelated* path dependence' (Martin and Sunley, 2006).

In reality, however, economies are typically ensembles of sectors in which productivity growth is linked by income and expenditure flows (Metcalfe et al., 2006). In a local or regional context there may well be interactions between industrial paths through, for example, upstream and downstream input–output linkages, knowledge spillovers, labour pooling, positive service and infrastructure externalities, or, conversely competition for land, finance and skilled labour. In other words, it is more likely that there is at least some degree of 'multiple *related* path dependence' across an urban or regional economy (Martin and Sunley, 2006). In a different context, Bassanini and Dosi (2001) argue that systems may be composed of path-dependent entities but whether or not the system as a whole will be path-dependent depends on the *structure of the interactions* between the constituent entities. They contend that when such interactions are strong

– above a certain 'threshold' – path dependence among the constituent entities induces path dependence at the level of the 'aggregate' system, although as they acknowledge, there is no a priori way of specifying what that 'interaction threshold' has to be before aggregate or macro-level path-dependent outcomes become observable. Grabher (1993), of course, argued that in the case of the Ruhr's iron and steel complex, strong inter-linkages had produced system-level (region-wide) path-dependent behaviour. In most regions there will be groups of inter-related or complementary industries and activities, linked either by direct input–output relationships, or by various indirect (or untraded) interdependencies and externalities. That is to say, various networks and structures of inter-relatedness can emerge between different sectors and activities within a region, thus suggesting the possibility of what we might call 'path-interdependence', that is situations where the path-dependent trajectories of particular local industries are to some degree interdependent. The extent and significance of this inter-linking path effect is a key issue for further research. Evolutionary economists increasingly emphasise how path dependence involves the *co-evolution* of different 'arenas' – such as the economic, technological, institutional and socio-cultural. And some geographers have shown how local path dependence in specialised clusters also occurs through such mutually reinforcing co-evolution of local economic, technological and socio-institutional 'arenas'. But path-dependent development may also involve the complex co-evolution of overlapping sets of inter-related industries, and thus may be just as important in shaping the evolution of economically diverse regions as it is in specialised ones. At present, however, there is a dearth of research into these regional co-evolutions so that we do not know how far they can effectively change the course of an industrial path, possibly by stimulating adaptation and rebalancing of the local industrial structure, once again disrupting any tendency towards a path dependent equilibrium state.

5. Economic evolution via new paths: place and path creation

In a David-type model of path-dependent equilibrium, the appearance of new paths is ascribed to the happenstance events or exogenous shocks to the system. But if path dependence can, in some circumstances, also be enabling then this gives us a much wider range of possible cases of new path creation (see Table 3.2). Path creation varies both in terms of whether past trajectories are enabling or constraining and in terms of whether origins are intended and deliberate or accidental. As Garud and Karnøe (2001) argue, we should not underestimate the role of purposeful, deliberate and strategic human action (also Schienstock, 2007). In their view, entrepreneurs are embedded in paths but not completely constrained by them; indeed they demonstrate 'mindful deviation' from prevailing ideas and accepted ways of doing things. However, it is important to realise that in most instances path creation will inevitably involve a complex admixture of deliberate agency and accidental and unintended emergence. Meyer and Schubert (2007), for example, argue that the formations of all technological paths lie somewhere on a continuum between completely accidental and random emergence, and deliberate intended creation. Agents are aware of emerging paths and invest in and bet on them, but the results of their actions are not always those intended. The constitution of new paths may be 'accidental' then in the sense that particular actions often have much longer and wider consequences than the actors intended or envisaged (David, 1999, Puffert, 2000). To return to Table 3.2, what we intend to show is that while the classic model of path

Table 3.2 Varieties of path creation

	Origins of new path of development	
Place and path effects	*Deliberate and intentional*	*Chance and accidental*
Enabling new paths	1. Agents search for opportunities, re-use resources, transfer competences as basis of new growth	2. Agents gain assets and experience, but accidents and events trigger new path
Constraining to existing path	3. Designed interventions to break path or switch location to overcome lock-in	4. Unpredictable external shocks and random events break old trajectory and launch new path

dependence is rooted in quadrant 4, more recent work in economic geography has begun to move closer to positions in quadrants 2 and 1, and has put much more emphasis on the re-use and transfer of resources and competences.

This wider set of possibilities raises a series of questions about the strength of path-dependent enabling effects. The dominant view in evolutionary economic geography has been that pre-existing paths of development are all but irrelevant in determining where new industrial paths emerge and become established. For example, Storper (1995) proposes the so-called 'window of locational opportunity' approach referred to above, in which – as he sees this idea – new technologies start as generic assets and then only subsequently evolve into specific assets. Initially, new sectors based on radically new technologies have few established specific inputs so that they 'invent' their own input chains and the associated knowledge, which is why, he claims, their initial location is basically serendipitous: this explains why, he argues, the semiconductor industry grew up in Silicon Valley and not alongside its parent industry – radio and television equipment – on the east coast. New industries enjoy moments of enhanced locational freedom called 'windows of locational opportunity' (Scott and Storper, 1987). In such periods, 'Capitalism is capable of escaping the past to create new localizations of industry' (Storper and Walker, 1989, p. 71). Similarly, according to Boschma and Frenken (2003, pp. 20–21).

> the evolutionary approach argues that the selection pressure of existing spatial structures is rather weak when new industries emerge. Under certain circumstances there are good reasons to assume that place specific features do not determine the location of new sectors. The environment is considered to be of minor importance at the initial stage of development of a sector when there exists a gap between the requirements of the new industry (in terms of knowledge, skills, etc.) and its surrounding environment. Windows of locational opportunity are open in emerging industries.

Thus, during the emergence of a new technology or industry, there will be a relatively large number of locations possessing the generic conditions that would allow the new sectors to grow there. New firms can locate where they please within this variety of places. What then happens, according to this thesis, is that one of these locational contenders is 'selected', often by inexplicable chance and random events, so that local enabling effects arising from previous patterns of economic development are very weak (Boschma, 2004; Boschma and Frenken, 2003; Krugman, 1991; Lambooy and Boschma, 2001). However, while this interpretation may have some validity in particular cases – and possibly in earlier periods of industrial history – the assumption that new paths require only generic assets that can subsequently be moulded is somewhat questionable as a general working hypothesis. Contemporary innovations combine inherited competences, experiences with customers and many fields of specialized knowledge (Pavitt, 2005). Moreover, this 'window of locational opportunity' view at times appears to conflate ex ante unpredictability with ex post inexplicability, and, as we will argue, while some types of selection pressure may be weak during the early phases of industry emergence, other types of selection process are critical and strong. As a consequence, this view underestimates the importance of place – of local economic and social history and context – in processes of path creation.

Places influence the origins of new technological-industrial paths in two main ways. First, place is likely to be implicated in processes of entrepreneurial variety generation, and second place is also likely to be important in processes of collective support, selection and the emergence of new trajectories. It is precisely the embedding of agents in particular paths that enables them to accumulate the resources and experience necessary to launch new technological paths (Garud and Karnøe, 2001). In this sense, the idea of *dependence between successive paths*, between successive industrial-technological trajectories *in particular places* becomes relevant. How exactly then might this occur?

The analysis of the emergence of high-technology clusters has emphasised the importance of pre-existing comparative advantages in shaping new increasing returns effects. As Bresnahan et al. (2005) argue, 'old economy' (or more generally, 'old' path) inputs, such as the supply of technical and managerial skills, connections to market niches, and the role of key firms, are often crucial in determining whether and where 'new economy' ('new' path) based increasing returns actually emerge and develop. While the development of technology clusters can be triggered by specific actions and events with unforeseen effects, these consequences only typically emerge because such events occur in the context of longer term conditions and facilitating factors. Thus while the growth of Silicon Valley was initiated by small contingent decisions such as Shockley's move to Palo Alto and the founding of Fairchild (Kenney and von Burg, 2001), there were also important preconditions including the development of the electronics industry in the area during the interwar period (Sturgeon, 2000). Another apt example is the Cambridge high-tech cluster (covering scientific instruments, software, and biotechnology, among other activities) in the UK. The origins of this are usually ascribed to the establishment of Cambridge Consultants in 1960 (a group of chemistry graduates concerned to foster research links between the university and local industry in the Cambridge area), or to the establishment of a science park by Trinity College in 1970. But they could equally well be traced back to antecedents in the 1930s, such as the existence of a specialist aeronautical instrument sector or the local agro-chemical industry.

In fact, there is increasing recognition of the ways in which resources and competences used in old paths may be recombined and reworked to form the basis of purposeful entrepreneurial deviations into new paths (Bathelt, 2001; Kemp et al., 2001; Metcalfe, 2005). A key mechanism here is the way in which spinoff firms inherit routines and competences from their parents. There is also mounting evidence that entrepreneurs transfer technological, business and marketing capabilities and competences from parent firms to their new ventures (Buenstorf, 2007; Klepper and Sleeper, 2005). Those spinoffs that inherit routines and competences from parent firms in the same or a related sector have often been found to be the most successful (Buenstorf and Klepper, 2008; Klepper, 2007). Where spinoffs enter a new but related sector, this may lead to path-dependent path creation, as spinoffs increase the diversity of economic knowledge and hence the potential for new products and technologies to emerge (Buenstorf and Fornahl, 2009; Klepper and Sleeper, 2005). It is well known that entrepreneurs tend to show locational inertia and prefer not to move home when starting a new venture, partly because of the importance of local social and business communities in mitigating risk (Stam, 2007). Related variety in a regional industrial structure is thus likely to be both an outcome of this transfer, and a resource that facilitates further diversification and innovation (Boschma and Frenken, 2006).

In an important sense, entrepreneurial opportunities are not exogenously given but are deliberately made. New ventures are based on 'business conceptions' or the entrepreneur's interpretation of the opportunity and the approach adopted to exploit it (Buenstorf, 2007; Witt, 1998). Such intuitive conceptions characteristically mix past experience, memory, and current intentions and they have important framing effects on the new firm's motivation and direction, and its capacity to discover new opportunities. The recognition of technological openings depends heavily on prior experience and learning (Shane, 2000). What is particularly important here is that, as Buenstorf (2007) notes, these conceptions help to pre-select among the variety of potential new activities. Ex ante selection is much more significant than implied by accounts that emphasise chance and 'accidental' events in path creation, as is illustrated by the example of technological niches.

Technological niches are usually shielded from some market selection pressures as they depend on fringe markets, experimental and special purpose users, or state subsidies and public research funding (Metcalfe et al., 2005; Schot and Geels, 2007). Ex post market selection may be weak in these contexts but, at the same time, such niches are typically marked by strong ex ante selection in which entrepreneurial agents imagine, conceive, design and gradually improve products and systems that they anticipate will meet demand or provide a basis for new growth. The critical moment here may be the application of existing technology to a new economic domain (Levinthal, 1998). Technological niches are typically marked by unstable technological rules and the lack of an established technological paradigm (Utterback, 1996). Their linkages are friable and fragile. Hence they exhibit a great deal of uncertainty and many information asymmetries and, in such environments, face-to-face contacts and local networks are important in building trust and knowledge exchange. Given that radical innovation is an inherently uncertain process, there will inevitably be a high failure rate in niche formation. New firms need to identify a fringe market that is not well served by old technologies, products or services, in order to keep themselves alive long enough to develop their own new technologies, products or services (Malerba et al., 2007). The test of a technological niche will depend on whether it can grow in small fringe market niches and then, ultimately, whether it can break through into a more general socio-technical regime (Geels, 2002, 2005; Geels and Schot, 2007). But we suspect that place plays an important role here as technologies, markets and selection pressures co-evolve.

A series of institutional and economic factors at local, regional and national scales can determine whether an innovation becomes the basis of a new industrial path or whether it remains isolated and underinvested and is unable to grow. Local institutions and human resources that have developed as a result of one industry's development in a region can also be critical causes of, and inputs to, the creation of other industries. These include the inherited entrepreneurial culture; social structures of innovation; access to specialist, demanding and knowledgeable customers; the presence of supporting institutions such as intermediaries, law and venture capital firms; and government provision of hard and soft infrastructures. Carlsson (2007, p. 265), for example, suggests that 'the most important aspect of path dependence may be the existing entrepreneurial climate resulting from pre-existing conditions'. Supporting social infrastructures, however, are difficult to orchestrate in a systematic fashion as they are collective and emergent. Typically, such institutions are not present at the birth of an industry but they gradually evolve as the local industry develops and as processes of positive lock-in consolidate the industrial path and reinforce its momentum (for example, Feldman, 2007). In addition, however, once a supportive generic institutional structure develops for one industry it can have beneficial consequences for subsequent newly emerging paths in other sectors (for example, see Zook, 2005). Thus while supportive local selection environments are typically secondary rather than primary causes of path creation, they can help to develop a technological niche or a radical innovation so that it stands a better chance of surviving market selection pressures.

In summary then, the equilibrium model of path dependence and its geographical offspring accord too little importance to the role of place in shaping new path creation. In this view, place is confined to a set of largely accidental and random initial conditions and triggers, which are followed by a set of cumulative and reinforcing increasing returns. Instead, we have argued that place dependence is important well before the unfolding of reinforcing dynamics as it conditions and influences the emergence of paths in particular sites. There is much evidence that local conditions continue to be important to processes of firm spinoff and to the emergence of radically new technological and innovation trajectories. While there is undoubtedly an unpredictable and uncertain dimension to path creation, this should not be exaggerated so as to completely obscure the deliberate ex ante selection of promising entrepreneurial ideas and the creative deployment of pre-existing resources, ideas and relationships. In recent work, chance is conceived as those random accidents and triggers that occur after the necessary accumulation of antecedent conditions and assets. But this dual-stage perspective can confuse micro-scale events with randomness, and its elaboration will need a much clearer exposition of contingency. Certainly, the place-specific path-dependent processes that help shape industrial development trajectories are far from deterministic and by no means easy to measure, as they will interact in a complex manner with extra-local contingencies and the basic unpredictability of radical innovation. It is not surprising, then, that so far there is very little research that explores these interacting effects, but it is no exaggeration to say that they are crucial to the long-run adaptability of urban and regional economies.

6. Conclusions: rethinking path dependence?

Our aim in this chapter has been to examine how far and in what ways path dependence can serve as an explicitly *evolutionary* concept for studying the economic landscape, thereby focusing on some of the issues that need to be addressed in any attempt to construct what Scott calls an 'ontology of regional growth and development that is rooted in the idea of path dependent economic evolution' (2006, p. 85).

Immediately, a central question has to be confronted: what sort of evolution – of a technology, and industry, or a local economy – is implied by the idea of path dependence? As articulated by its leading exponent, Paul David, the conception is one in which the contingencies of historical accident act to select between multiple (and perhaps competing) possibilities and then various feedback forces come into operation that serve to reproduce the selected particular technology or industry, a situation that he argues can be thought of as the contingent selection of, and lock-in to, one of a number of possible multiple equilibrium states. In his model it takes an exogenous shock to disrupt that equilibrium state, so that economic evolution takes the form of successive punctuated equilibria. David refers to this conception as 'strong history' (see also Castaldi and Dosi, 2004).

We have argued, however, that this conception is overly restrictive, too 'strong' we might say, and that David's very notion of 'path-dependent equilibrium economics' is itself something of a contradiction in terms. Of course there are examples of this sort of selection and lock-in of a technology and even an industry to a stable, self-reproducing state; but even such instances need not represent equilibrium situations. According to complexity theory, for example, it is possible for a system to exhibit stability or inertia even though it is 'far from equilibrium'. Moreover, examples of technologies, industries – let alone whole local or regional economies – being in a stable and self-reproducing unchanging state are not the norm. Most technologies, industries and local and regional economies follow development paths that evolve over time, in a path-dependent manner. The idea of multiple equilibria, to our mind, fails to capture this process. We have sought to argue, therefore, for a richer interpretation of path-dependent economic evolution, one that does not require or necessitate notions of equilibrium.

Instead, we conceive of the idea of path dependence as entirely consistent with patterns of economic evolution in which technologies, industries, institutions, and regional economies adapt and mutate over time without ever reaching or tending towards any equilibrium. David, no doubt, would criticise this view as one of 'weak history' and argue that it is not, therefore, path dependence. But, arguably, it is a history that accords with much of observable reality. Whereas in David's view of path dependence, history to all intents and purposes ceases once the system in question becomes locked into its 'equilibrium state', in our conception history continues to unfold, in a path-dependent manner. This means that the focus of analysis centres on how economic evolution takes place along paths, and on whether and in what ways such paths undergo 'life-cycles', and on how old paths are replaced by new.

In fact, in searching for this sort of path dependence ontology, we believe economic geographers could benefit from looking outside the interpretation found within economics to the wider discussion of path dependence and historical change that has taken place in recent years, not only across the social, political and policy sciences, but also in the field of ecological studies.[20] As Mahoney (2000, 2006) and Howlett and Rayner (2006) argue, path dependence is but one of a number of possible models of historical change for socio-economic contexts, and social scientists have introduced important modifications to the idea of path dependence that go beyond what the economics discipline currently offers. For example, Mahoney (2000, 2006) suggests that a somewhat different conception of path dependence can be constructed in terms of what he has called 'reactive sequences'. Reactive sequences are chains of temporally ordered and causally ordered events. In a reactive sequence, each event in the historical sequence is both a reaction to antecedent events and a causal influence on subsequent events. This still means that early, initial events are important to later outcomes: small changes in initial conditions can accumulate over time and make a great deal of difference by the end of a sequence. In this respect, reactive sequences retain one of the key aspects of the path dependence idea. But in other respects, reactive sequences are quite different from the basic 'lock-in to equilibrium' conception of David-type path dependence models. Whereas the increasing returns sequences that are typically invoked in standard path dependence models are characterised by processes of reproduction that reinforce early events, reactive sequences are marked by processes that transform earlier events: in a reactive sequence, initial events trigger subsequent development not by reproducing a given pattern, but by setting in motion a chain of linked reactions and events. As such, this interpretation of path dependence offers a much more open set of possibilities in terms of the trajectories of technological, industrial and regional economic paths.

Another possible framework is the so-called 'process-sequencing' model (Howlett and Rayner, 2006). This conceives of social processes as the connections between events in different time periods arising in terms of reiterated problem solving (Haydu, 1998). Proponents of this model among sociologists argue that it has advantages over standard path dependency models as it 'provides a plausible way to represent and account for historical trajectories; it builds social actors and multiple causal timelines into explanatory accounts; and it offers a richer sense of how earlier outcomes shape later ones' (Haydu, 1998, p. 341). That is, 'continuities across temporal cases can be traced in part to enduring problems, while more or less contingent solutions to those problems are seen as reflecting and regenerating the historical individuality of each period' (Haydu, 1998, p. 354). Unlike path dependency, process sequencing emphasises not how outcomes at historical switch points are accidents, but how they are firmly based or rooted in previous events and developments as related structural processes of negative and positive feedback that affect actor behaviour (Baumgartner and Jones, 2002). Changes in trajectories in this model are not random or chaotic, but are outgrowths of earlier trajectories. Hence, although process sequencing shares some of the characteristics of the standard path dependency model, it is not the same.

Yet another approach worthy of study by economic geographers is the so-called 'panarchy' perspective that has been proposed in evolutionary ecology (see, for example, Gunderson and Holling, 2002). Panarchy is a conceptual framework that seeks to account for the seemingly contradictory characteristic of all complex systems – persistence and change. Great emphasis is put on the interconnectedness of levels, between the smallest and the largest, and the fastest and slowest. Large scale slow processes set the conditions for the smaller scale faster processes to operate. But the small scale fast changes can also have an impact on the larger, slower processes. The fast levels invent, experiment and test; the slower levels stabilise and conserve the accumulated memory of past, successful experiments. The primary focus of this approach is on the nature and dynamics of 'adaptive cycles', the process that accounts for both stability and change. The adaptive cycle periodically generates variability and novelty, either internally through mutation and adaptation or by accumulating resources that change the internal dynamics of the system. The issue of system 'resilience' also figures prominently in this approach. While this framework is still in its infancy, and thus far has been primarily ecological in focus, its core ideas have been argued to be applicable to a wide range of social, political and economic contexts. Certain of its notions resonate with those from the theory of complex adaptive systems, and like the latter it is concerned with processes not unrelated to those of path dependence. The emphasis on different scales and speeds of change, and their interaction, would seem of particular potential relevance to the sort of open, non-equilibrium conception of path dependence that we have argued for here.

If the objective of a path dependence approach to understanding regional growth and development is to uncover and make sense of the specific historical mechanisms and chains of events that have produced a particular economic landscape (rather than some other possible configuration), then the idea of path dependence should not, in our view, be circumscribed by chaining it to any form of equilibrium thinking. To do so is to severely restrict its potential as an evolutionary concept. A rethink of the notion of path dependence may well be required to ensure it does not fall prey to 'intellectual lock-in' that seems to bind so much of economics – even some versions of evolutionary economics – to equilibrium thinking.

Notes

1. This contribution builds on our earlier paper, Martin and Sunley (2006). Ron Martin's research for this chapter was undertaken while holding a Leverhulme Major Research Fellowship.
2. In economics, ideas akin to 'path dependence' can in fact be traced back more than a century, to Carl Menger's (1883 [1985]) analysis of 'institutional emergence' and Thorstein Veblen's (1898) concept of 'cumulative causation' in the evolution of habits and conventions. And the closely related concept of 'hysteresis' has also been around since the mid-1970s (see Cross, 1993; Elster, 1976; Franz, 1990; Katzner, 1993). Outside of economics we find the concept being applied to topics as diverse as decision-making and social behaviour (Anderlini and Ianni, 1996; Goldstone, 1998); industrial organisation (Antonelli, 1997); power generation technologies (Cowan, 1990); pest control programmes (Cowan and Gunby, 1996); industrial technology strategies (Araujo and Harrison, 2002; Ruttan, 1997); technological leadership (Redding, 2002); corporate governance (Bebchuk and Roe, 1999); legal systems and social institutions (North, 1990); historical sociology (Goldstone, 1998; Mahoney, 2000); corporate organisation (Sydow et al., 2005); and politics and state intervention (Bridges, 2000; Dimittakopoulos, 2001; Magnusson, 2001; Pierson, 2000). This proliferation of path dependence ideas can itself be interpreted as a reflection of what some have referred to as a 'historical turn' in the social and cognate sciences (Abbott, 2001; Howlett and Rayner, 2006; Mahoney and Rueschemeyer, 2003).
3. This logic had in fact been anticipated somewhat earlier by Storper and Walker who argued that 'Localized technological change in an industry can be understood, like all industrial development, as an evolutionary path in which each step moves one way from a past that cannot be recovered and that limits future directions' (1989, p. 113).
4. In focusing on the role and meaning of path dependence as an evolutionary concept, we by no means wish to imply that the other issues concerning the notion of path dependence set out in Table 3.1 are secondary or unimportant: far from it. But arguably the role of path dependence as an evolutionary concept is the most fundamental. For a discussion of the other issues raised in Table 3.1, see Martin and Sunley (2006).
5. While there are many different definitions of such an equilibrium, according to Setterfield (1997), an equilibrist methodology of this sort is typically characterised by two distinctive features: first, the specification of a model of structural equations conditioning endogenous variables on exogenously given 'data' – usually a set of variables and coefficients whose values are imposed on the system from without; and, second, the construction of such models so as to yield stable equilibria, that is points to which the system will return following some initial displacement.
6. The historical sequence, $h(t)x$, could also include the past values of other external factors, say y, that have also shaped the development path of x.
7. For example, the same seemingly irresistible attachment to equilibrium is to be found in Page's (2006)

otherwise very useful attempt to distinguish between different forms and degrees of path dependence. He states: 'of course a process need not attain an equilibrium distribution or outcome, but for the purposes of this essay I restrict attention to that case [of equilibrium], with one exception' (p. 92).

8. More explicitly, David defines path dependence in terms of non-ergodic Markov chain models that possess two or more absorbing states. An absorbing state is one from which there is a zero probability of exit or movement to another state. In such models, the limiting distribution depends on the initial distribution, that is on where the system started: different starting distributions will result in different limiting (equilibrium) distributions: hence the idea of multiple equilibria. We should note here, however, anticipating the discussion that follows, that absorbing Markov chains are only one form of non-ergodic process, and that in general the latter do not have to possess any long-term limiting distribution or stable equilibrium state.
9. The literature on NEG models is now extensive, but for representative expositions see Fujita et al. (1999), Baldwin et al. (2003), Henderson (2005) or Brakman et al. (2009). For a recent assessment of these models from a 'proper' economic geography perspective, see Martin (2010a).
10. Although Setterfield makes this argument in the context of deterministic systems, it applies equally to stochastic systems like the absorbing Markov chain processes that David uses to conceptualise path dependence.
11. Of course, it could be argued that equilibrium economics has long moved on beyond traditional equilibrist methodology, and while retaining its focus on the existence and achievement of equilibrium, has come to embrace the idea that history matters through the construction of models of sequence economics, temporary equilibrium, and non-tatonnement processes. But as Setterfield retorts, these extensions actually constitute 'nothing more than glorified conventional disequilibrium analysis and are roughly in keeping with equilibrist methodology . . . the language of the analysis may be that of sequences and temporary equilibria, but the substance does not involve path dependency, as the economy is held ultimately to converge to a predefined, "fully adjusted" equilibrium position' (Setterfield, 1997, p. 62).
12. Bassinini and Dosi's discussion falls victim to this last problem when they seem to invoke radical technological shifts – between steam power and electricity, and between thermionic valves and semi-conductors – as examples of 'temporary' resting states.
13. As Schumpeter (1944) insisted, transformation arises from *within* the socio-economic system, and adaptive development is the primary process by which this occurs.
14. Ramlogan and Metcalfe make the point that if knowledge ever did reach an equilibrium state then all economic change and development would cease.
15. Complexity theorists usually use the term hysteresis rather than path dependence. Though not identical, the two notions are related.
16. The rise of agent-based models within some quarters of neoclassical economics can be argued to be a manifestation of recognising the value of assigning more importance to economic process than to end states (see, for example, Axtell, 2007).
17. We prefer the phrase 'non-equilibrium' here, rather than 'disequilibrium' or 'out-of-equilibrium', since both of the latter continue to imply that there is some latent equilibrium outcome to which path dependence would lead if only the system in question were free to do so. The term 'non-equilibrium' is intended to imply that no such latent equilibrium may exist at all.
18. It is analogous in a sense to Paul Davidson's (1991) post-Keynesian economics in which decision-making is so non-ergodic that the past is no real guide to the future.
19. In evolutionary economics, adaptation refers to a process of response to the selection environment so that adapted entities fit the environment and are apt or fit for purpose (Metcalfe, 2005). Adaptability, on the other hand, refers to 'the potential to adjust to changing circumstances in an appropriate way: it is about the capacity to respond to changes in the selection environment: to maintain good design' (Metcalfe, 2005, p. 414). However, economic adaptation remains a layered and contested concept: does it refer, for instance, to the process in which entities ensure that they are fit for market survival or to the process of using resources appropriately and satisfying needs, as clearly the two may overlap but are not synonymous? It is important therefore to clarify the precise firm strategy taken in response to a specified environment and to examine its implications for adaptability.
20. The exploration of alternative approaches to path dependence being developed in political science and historical sociology is taken up in more detail in Martin (2010b).

References

Abbott, A. (2001), *Time Matters: On Theory and Method*, Chicago: University of Chicago Press.

Agarwal, R. (1998), 'Evolutionary trends of industry variables', *International Journal of Industrial Organization*, **16** (4), 511–25.

Agarwal, R. and M. Gort (1996), 'The evolution of markets and entry, exit and survival of firms', *Review of Economics and Statistics*, **78** (3), 489–98.

Anderlini, L. and A. Ianni (1996), 'Path dependence and learning from neighbours', *Games and Economic Behaviour*, **13**, 141–77.
Antonelli, C. (1997), 'The economics of path dependence in industrial organisation', *International Journal of Industrial Organisation*, **15** (6), 643–75.
Araujo, L. and D. Harrison (2002), 'Path dependence, agency and technological evolution', *Technology Analysis and Strategic Management*, **14**, 5–19.
Arthur, W.B. (1988), 'Self-reinforcing mechanisms in economics', in P. Anderson, K. Arrow and D. Pines (eds), *The Economy as an Evolving, Complex System*, Reading, MA: Addison-Wesley, pp. 9–31.
Arthur, W.B. (1989), 'Competing technologies, increasing returns, and "lock-in" by historical events', *Economic Journal*, **99**, 116–31.
Arthur, W.B. (1994a), *Increasing Returns and Path Dependence in the Economy*, Michigan: Michigan University Press.
Arthur, W.B. (1994b), 'Path dependence, self-reinforcement and human learning', in *Increasing Returns and Path Dependence in the Economy*, Michigan: Michigan University Press, pp. 133–58.
Arthur, W.B. (1994c), 'Industry location patterns and the importance of history', *Increasing Returns and Path Dependence in the Economy*, Michigan: Michigan University Press, pp. 49–67.
Arthur, W.B. (1994d), 'Urban systems and historical path dependence', in *Increasing Returns and Path Dependence in the Economy*, Michigan: Michigan University Press, pp. 99–110.
Audretsch, D.B. and M.P. Feldman (1996), 'R&D spillovers and the geography of innovation and production', *The American Economic Review*, **86** (3), 630–40.
Axtell, R.L. (2007), 'What economic agents do: how cognition and interaction lead to emergence and complexity', *Review of Austrian Economics*, **20** 105–22.
Baldwin, R., R. Forslid, P. Martin, G. Ottaviano and F. Robert-Nicoud (2003), *Economic Geography and Public Policy*, Princeton, NJ: Princeton University Press.
Bassanini, A.P. and G. Dosi (2001), 'When and how chance and human will can twist the arms of Clio: an essay on path dependence in a world of irreversibilities', in R. Garud and P. Karnøe (eds), *Path Dependence and Creation*, London: Lawrence Erlbaum, pp. 41–68.
Bathelt, H. (2001), 'Regional competence and economic recovery: divergent growth paths in Boston's high technology economy', *Entrepreneurship and Regional Development*, **13**, 287–314.
Bathelt, H. and J.S. Boggs (2003), 'Towards a reconceptualization of regional development paths: is Leipzig's media cluster a continuation of or a rupture with the past?, *Economic Geography*, **79** (2), 265–93.
Baumgartner, F. and B. Jones (eds) (2002), *Policy Dynamics*, Chicago: University of Chicago Press.
Bebchuk, A. and M. Roe (1999), 'A theory of path dependence in corporate ownership and governance', *Stanford Law Review*, **52**, 127–70.
Beinhocker, E.D. (2006), *The Origin of Wealth: Evolution, Complexity and the Radical Remaking of Economics*, Cambridge, MA: Harvard Business School Press.
Bode, E. (2001), 'Is regional innovative activity path dependent? An empirical analysis for Germany', *Working Paper 1058*, Kiel Institute for World Economics, Kiel.
Boschma, R.A. (2004), 'Competitiveness of regions from an evolutionary perspective', *Regional Studies*, **38** (9), 1001–14.
Boschma, R. (2005), 'Proximity and innovation: a critical assessment', *Regional Studies*, **39** (1), 61–74.
Boschma, R.A. and K. Frenken (2003), 'Evolutionary economics and industry location', *International Review for Regional Research*, **23**, 183–200.
Boschma, R.A. and K. Frenken (2006), 'Why is economic geography not an evolutionary science?', *Journal of Economic Geography*, **6**, 273–302.
Boschma. R. and B. van der Knaap (1997), 'New technology and windows of locational opportunity', in J. Reijnders (ed.), *Economics and Evolution*, Cheltenham, UK and Lyme, NH, USA: Edward Elgar, pp. 171–202.
Brakman, S., H. Garretsen and C. Marrewijk (2009), *The New Introduction to Geographical Economics*, Cambridge: Cambridge University Press.
Bresnahan, T., A. Gambardella and A. Saxenian (2005), '"Old economy" inputs for "new economy" outcomes: cluster formation in the new Silicon Valleys', in S. Breschi and F. Malerba (eds), *Clusters, Networks and Innovation*, Oxford: Oxford University Press, pp. 113–35.
Bridges, A. (2000), 'Path dependence, sequence, history, theory', *Studies in American Political Development*, **14** (1), 109–12.
Buenstorf, G. (2007), 'Evolution on the shoulders of giants: entrepreneurship and firm survival in the German laser industry', *Review of Industrial Organization*, **30** (3), 179–202.
Buenstorf, G. and D. Fornahl (2009), 'B2C – bubble to cluster: the dot-com boom, spinoffs, entrepreneurship and regional agglomeration', *Journal of Evolutionary Economics*, **19** (3), 349–78.
Buenstorf, G. and S. Klepper (2008), 'Heritage and agglomeration: the Akron tyre cluster revisited', *The Economic Journal*, **119** (537), 705–37.

Carlsson, B. (2007), 'The role of public policy in emerging clusters', in P. Braunerhjelm and M. Feldman (eds), *Cluster Genesis: Technology-Based Industrial Development*, Oxford: Oxford University Press, pp. 264–78.

Castaldi, C. and G. Dosi (2004), 'The grip of history and the scope for novelty: some results and open questions on path dependence in economic processes', *Working Paper 2003/02*, Laboratory of Economics and Management, Sant'Anna School of Advanced Studies, University of Pisa.

Castellacci, F. (2006), 'A critical realist interpretation of evolutionary growth theorising', *Cambridge Journal of Economics*, **30** (6), 861–80.

Chapman, K. (2005), 'From growth centre to cluster: restructuring, regional development, and the Teeside chemical industry', *Environment and Planning A*, **37**, 597–615.

Chapman, K., D. MacKinnon and A. Cumbers (2004), 'Adjustment or renewal in regional clusters: a study of networking among SMEs in the Aberdeen oil complex', *Transactions of the Institute of British Geographers*, **29**, 382–96.

Cooke. P. and K. Morgan (1998), *The Associational Economy*, Oxford: Oxford University Press.

Cowan, R. (1990), 'Nuclear power reactors: a study in technological "lock-in"' *Journal of Economic History*, **50** (3), 541–67.

Cowan, R. and P. Gunby (1996), 'Sprayed to death: path dependence, "lock-in" and pest control strategies', *Economic Journal*, **106**, 521–42.

Cross, R. (1993), 'On the foundations of hysteresis in economic systems', *Economics and Philosophy*, **9**, 53–74.

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

David, P.A. (1986), 'Understanding the economics of QWERTY: the necessity of history', in W.N. Parket (ed.), *Economic History and the Modern Economics*, Oxford: Blackwell.

David, P.A. (1988), 'Path dependence: putting the past into the future of economics', *The Economic Series Technical Report 533*, Institute for Mathematical Studies in the Social Sciences, Stanford University, California.

David, P.A. (1992), 'Heroes, herds and hysteresis in technological history: Thomas Edison and "The Battle of the Systems" reconsidered', *Industrial and Corporate Change*, **1** (1), 129–80.

David, P.A. (1993a), 'Path dependence and predictability in dynamic systems with local network externalities: a paradigm for historical economics', in D. Foray and C. Freeman (eds), *Technology and the Wealth of Nations*, London: Pinter.

David, P.A. (1993b), 'Historical economics in the long run: some implications of path dependence', in G.D. Snooks (ed.), *Historical Analysis in Economics*, London: Routledge.

David, P.A. (1994), 'Why are institutions the "carriers of history"? Path dependence and the evolution of conventions, organisations and institutions', *Structural Change and Economic Dynamics*, **5** (2), 205–20.

David, P.A. (1999), 'At last: a remedy for chronic QWERTY-skepticism', *Discussion Paper for the European Summer School in Industrial Dynamics*, Institute d'Etudes Scientifiques de Cargese, September.

David, P.A. (2001), 'Path dependence, its critics and the quest for "historical economics"', in P. Garrouste and S. Ioannides (eds), *Evolution and Path Dependence in Economic Ideas*, Cheltenham, UK and Northampton, MA, USA: Edward Elgar, pp. 15–40.

David, P.A. (2005a), 'Path dependence in economic processes: implications for policy analysis in dynamical systems contexts', in K. Dopfer (ed.), *The Evolutionary Foundations of Economics*, Cambridge: Cambridge University Press, pp. 151–94.

David, P.A. (2005b), 'Path dependence and historical social science: an introductory lecture', Paper presented at the Symposium on Twenty Years of Path Dependence and Qwerty Effects, SIEPR Policy Paper No. 04–022, Russian University Higher School of Economics, Moscow, 13 May 2005.

David, P.A. (2007), 'Path dependence and historical social science: an introductory lecture, Paper presented at the *Symposium on Twenty Years of Path Dependence and Qwerty-Effects*, Russian University Higher School of Economics, Moscow, 13 May 2005, SIEPR Policy Paper No. 04–022.

Davidson, P. (1991), 'Is probability theory relevant for uncertainty? A post-Keynesian perspective', *Journal of Economic Perspectives*, **5** (1), 129–43.

Dimitrakopoulos, D.G. (2001), 'Incrementalism and path dependence', *Journal of Common Market Studies*, **39**, 405–22.

Durlauf, S. (1997), 'Limits to science or limits to epistemology', *Complexity*, **2** (3), 31–7.

Elster, J. (1976), 'A note on hysteresis in the social sciences', *Synthese*, **33**, 371–91.

Essletzbichler, J. and D. Rigby (2007), 'Exploring evolutionary economic geographies', *Journal of Economic Geography*, **5**, 549–71.

Feldman, M.P. (2007), 'Perspectives on entrepreneurship and cluster formation: biotechnology in the US Capitol region', in K. Polenske (ed.), *The Economic Geography of Innovation*, Cambridge: Cambridge University Press, pp. 241–60.

Feldman, M.P. (2005), 'The entrepreneurial event revisited: firm formation in a regional context', in S. Breschi and F. Malerba (eds), *Clusters, Networks and Innovation*, Oxford: Oxford University Press, pp. 136–68.

Franz, W. (1990), 'Hysteresis in economic relationships: an overview', *Empirical Economics*, **15**, 109–25.

Fuchs, G. and P. Shapira (eds) (2005), *Rethinking Regional Innovation and Change: Path Dependency or Regional Breakthrough?*, New York: Springer Verlag, pp. 23–42.

Fujita, M. and T. Mori (2005), 'Frontiers of the new economic geography', *Papers in Regional Science*, **84** (3), 377–405.

Fujita, M., P. Krugman and T. Venables (1999), *The Spatial Economy*, Cambridge, MA: MIT Press.

Garud, R. and P. Karnøe (2001), 'Path creation as a process of mindful deviation', in R. Garud and P. Karnøe (eds), *Path Dependence and Creation*, London: Lawrence Erlbaum, pp. 1–38.

Geels, F. (2002), 'Technological transitions as evolutionary reconfiguration processes: a multi-level perspective and a case-study', *Research Policy*, **31**, 1257–74.

Geels, F. (2005), 'Processes and patterns in transitions and systems innovations: refining the co-evolutionary multi-level perspective', *Technological Forecasting and Social Change*, **72**, 681–96.

Geels, F. and J. Schot (2007), 'Typology of sociotechnical transition pathways', *Research Policy*, **36**, 399–417.

Gertler, M. (2005), 'Tacit knowledge, path dependency and local trajectories of growth', in G. Fuchs and P. Shapira (eds), *Rethinking Regional Innovation and Change: Path Dependency or Regional Breakthrough?*, New York: Springer Verlag, pp. 23–42.

Glasmeier, A. (2000), *Manufacturing Time: Global Competition in the Watch Industry, 1795–2000*, New York: Guilford Press.

Glasmeier, A. (2007), 'The self-aware firm: information needs, acquisition strategies, and utilization prospects', in K. Polenske (ed.), *The Economic Geography of Innovation*, Cambridge: Cambridge University Press, pp. 112–28.

Goldstone, J.A. (1998), 'Initial conditions, general laws, path dependence and explanation in historical sociology', *American Journal of Sociology*, **104** (3), 829–45.

Grabher, G. (1993), 'The weakness of strong ties: the "lock-in" of regional development in the Ruhr area', in G. Grabher (ed.), *The Embedded Firm: On the Socio-Economics of Industrial Networks*, London: Routledge.

Gunderson, L. and C.S. Holling (eds) (2002), *Panarchy: Understanding Transformations in Human and Natural Systems*, Washington, DC: Island Press.

Hall, P. (1994), *Innovation, Economics and Evolution: Theoretical Perspective on Changing Technology in Economic Systems*, London; Harvester Wheatsheaf.

Harris, D.J. (2004), 'Joan Robinson on "history versus equilibrium"', Paper presented at the Joan Robinson Centennial Conference, Department of Economics, University of Vermont, Burlington, 17–19 October 2003.

Harvey, D. (1982), *Limits to Capital*, Oxford: Basil Blackwell.

Harvey, D. (1985), 'The geopolitics of capitalism', in D. Gregory and J. Urry (eds), *Social Relations and Spatial Structures*, London: Macmillan, pp. 128–63.

Harvey, D. (2006), *Spaces of Global Capitalism: Towards a Theory of Uneven Geographical Development*, London: Verso.

Hassink, R. (2005a), 'How to unlock regional economies from path dependency? From learning region to learning cluster', *European Planning Studies*, **13** (4), 521–35.

Hassink, R. (2005b), 'Geography, networks and renewal in old industrial areas', Paper prepared for Fourth European Meeting on Applied Evolutionary Economics, Utrecht, 19–21 May 2005.

Haydu, J. (1998), 'Making use of the past: time periods as cases to compare and as sequences of problem solving', *American Journal of Sociology*, **104** (2), 339–71.

Heffernan, G. (2003), 'Path dependence, behavioural rules and the role of entrepreneurship in economic change: the case of the automobile industry', *Review of Austrian Economics*, **16** (1), 45–62.

Henderson, J.V. (ed.) (2005), *New Economic Geography*, Cheltenham, UK and Northampton, MA, USA: Edward Elgar.

Howlett, M. and J. Rayner (2006), 'Understanding the historical turn in the policy sciences: a critique of stochastic, narrative, path dependency and process-sequencing models of policy-making over time', *Policy Science*, **39**, 1–18.

Hudson, R. (2005), 'Rethinking change in old industrial regions: reflecting on the experience of North East England', *Environment and Planning A*, **37**, 581–96.

Kaldor, N. (1934), 'A classificatory note on the determinateness of equilibrium', *Review of Economic Studies*, **1** (2), 122–36.

Katzner, D. (1993), 'Some notes on the role of history and the definition of hysteresis and related concepts in economic analysis', *Journal of Post-Keynesian Economics*, **15**, 323–45.

Kaldor, N. (1972), 'The irrelevance of equilibrium economics', *The Economic Journal*, **82** (328), 1237–55.

Kaldor, N. (1985), *Economics Without Equilibrium*, Cardiff: University College of Cardiff Press.

Kemp, R., A. Rip and J. Schot (2001), 'Constructing transition paths through the management of niches', in R. Garud and P. Karnøe (eds), *Path Dependence and Creation*, Mahwah, NJ: Lawrence Erlbaum Associates, pp. 269–99.

Kenney, M. and U. von Burg (2001), 'Paths and regions: the creation and growth of Silicon Valley', in R. Garud and P. Karnøe (eds), *Path Dependence and Creation*, London: Lawrence Erlbaum, pp. 127–48.

Klepper, S. (1996), 'Entry, exit, growth, and innovation over the product life-cycle', *The American Economic Review*, **86** (3), 562–83.
Klepper, S. (1997), 'Industry life cycles', *Industrial and Corporate Change*, **6** (1), 145–82.
Klepper, S. (2002), 'Firm survival and the evolution of oligopoly', *The Rand Journal of Economics*, **33** (1), 37–61.
Klepper, S. (2007), 'Disagreements, spinoffs and the evolution of Detroit as the capital of the US auto industry', *Management Science*, **53** (4), 616–31.
Klepper, S. and K.L. Simons (2005), 'Industry shakeouts and technological change', *International Journal of Industrial Organization*, **23** (1–2), 23–43.
Klepper, S. and S. Sleeper (2005), 'Entry by spinoffs', *Management Science*, **51** (8), 1291–306.
Klepper, S. and P. Thompson (2006), 'Submarkets and the evolution of market structure', *Rand Journal of Economics*, **37** (4), 861–86.
Krugman, P. (1991), *Geography and Trade*, Leuven and Cambridge, MA: Leuven University Press and MIT press.
Lambooy J.G. and R.A. Boschma (2001), 'Evolutionary economics and regional policy', *Annals of Regional Science*, **35**, 113–31.
Levinthal, D. (1998), 'The slow pace of rapid technological change: gradualism and punctuation in technological change', *Industrial and Corporate Change*, **7** (2), 217–47.
Magnusson, L. (2001), 'The role of path dependence in the history of regulation', in L. Magnusson and J. Ottosson (eds), *The State, Regulation, and the Economy: An Historical Perspective*, Cheltenham, UK and Northampton, MA, USA: Edward Elgar.
Mahoney, J. (2000), 'Path dependence in historical sociology', *Theory and Society*, **29** (4), 507–48.
Mahoney, J. (2006), 'Analyzing path dependence: lessons from the social sciences', in A. Wimmer and R. Kössler (eds), *Understanding Change: Models, Methodologies and Metaphors*, London: Palgrave Macmillan, pp. 129–39.
Mahoney, J. and D. Rueschemeyer (2003), *Comparative Historical Analysis in the Social Sciences*, Cambridge: Cambridge University Press.
Malerba, F., R. Nelson, L. Orsenigo and S. Winter (1999), 'History-friendly models of industry evolution: the computer industry', *Industrial and Corporate Change*, **8** (1), 3–40.
Malerba, F., R. Nelson, L. Orsenigo and S. Winter (2007), 'Demand, innovation, and the dynamics of market structure: the role of experimental users and diverse preferences', *Journal of Evolutionary Economics*, **17**, 371–99.
Markusen, A. (1985), *Profit Cycles, Oligopoly and Regional Development*, Cambridge, MA: MIT Press.
Martin, R.L. (2010a), 'The new economic geography: credible models of the economic landscape?', in R. Lee, A. Leyshon, L. McDowell and P. Sunley (eds), *A Compendium of Economic Geography*, London: Sage.
Martin, R.L. (2010b), 'The Roepke lecture in economic geography: rethinking regional path dependence: from lock-in to evolution', *Economic Geography*, **86** (1), 1–27.
Martin, R.L. and P. Sunley (2006), 'Path dependence and regional economic evolution', *Journal of Economic Geography*, **6**, 395–438.
Martin, R.L. and P. Sunley (2007), 'Complexity thinking and evolutionary economic geography', *Journal of Economic Geography*, **7** (5), 573–601.
Massey, D. (1984), *Spatial Divisions of Labour*, London: Macmillan.
Menger, C. (1883) (1985), *Investigation into the Method of the Social Sciences*, New York: New York University Press.
Menzel, T. and D. Fornahl, D. (2007), 'Cluster life cycles – dimensions and rationales of cluster evolution', Jena Economic Research Paper 2007–076.
Metcalfe, J.S. (2005), 'Evolutionary concepts in relation to evolutionary economics', in K. Dopfer (ed.), *The Evolutionary Foundations of Economics*, Cambridge: Cambridge University Press, pp. 391–430.
Metcalfe, J.S., J. Foster and R. Ramlogan (2006), 'Adaptive economic growth', *Cambridge Journal of Economics*, **30**, 7–32.
Metcalfe, J., A. James and A. Mina (2005), 'Emergent innovation systems and the delivery of clinical services: the case of intra-ocular lenses', *Research Policy*, **34**, 1283–304.
Meyer, U. and C. Schubert (2007), 'Integrating path dependency and path creation in a general understanding of path constitution: the role of agency and institutions in the stabilisation of technological innovations', *Science, Technology and Innovation Studies*, **3**, 23–44.
North, D.C. (1990), *Institutional Change and Economic Performance*, Cambridge: Cambridge University Press.
Page, S.E. (2006), 'Path dependence', *Quarterly Journal of Political Science*, **1**, 87–115.
Pavitt, K. (2005), 'Innovation processes', in J. Fagerberg, D.C. Mowery and R.R. Nelson (eds), *The Oxford Handbook of Innovation*, Oxford: Oxford University Press, pp. 86–114.
Pierson, P. (2000), 'Increasing returns, path dependence and the study of politics', *American Political Science Review*, **94** (2), 252–67.

Puffert, D. (2000), 'Path dependence, network form and technological change', Paper presented at the Conference to Honour Paul David – History Matters: Economic Growth, Technology, and Population, Stanford University.
Ramlogan, R. and J.S. Metcalfe (2006), 'Restless capitalism: a complexity perspective on modern capitalist economies', in E. Garnsey and J. McGlade (eds), *Complexity and Co-evolution: Continuity and Change in Socio-Economic Systems*, Cheltenham: Harper, pp. 115–46.
Rantisi, N. (2002), 'The local innovation system as a source of "Variety": openness and adaptability in New York City's garment district', *Regional Studies*, **36** (6), 587–602.
Redding, S. (2002), 'Path dependence, endogenous innovation and growth', *International Economic Review*, **43**, 1215–48.
Robinson, J. (1974), *History Versus Equilibrium,* Thames Papers in Political Economy, Thames Polytechnic, London.
Ruttan, V.W. (1997), 'Induced innovation, evolutionary theory and path dependence: sources of technical change', *Economic Journal*, **107** (4), 1520–29.
Saxenian, A. (1996), *Regional Advantage: Culture and Competition in Silicon Valley and Route 128*, Cambridge, MA: Harvard University Press.
Schienstock, G. (2007), 'From path dependency to path creation: Finland on the way to the knowledge economy', *Current Sociology*, **55** (1), 92–109.
Schot, J. and F. Geels (2007), 'Niches in evolutionary theories of technical change', *Journal of Evolutionary Economics*, **17**, 605–22.
Schumpeter, J. (1944), *Capitalism, Socialism and Democracy*, London: George Allen & Unwin.
Scott, A.J. (2006), *Geography and Economy*, Oxford: Oxford University Press.
Scott, A.J. and M. Storper (1987), 'High technology industry and regional development: a theoretical critique and reconstruction', *International Social Science Journal*, **1** (12), 215–32.
Setterfield, M. (1995), 'Historical time and economic theory', *Review of Political Economy*, **7**, 1–27.
Setterfield, M. (1997), 'Should economists dispense with the notion of equilibrium?', *Journal of Keynesian Economics*, **20** (1), 47–76.
Shane, S. (2000), 'Prior knowledge and the discovery of entrepreneurial opportunities', *Organization Science*, **11** (4), 448–69.
Smith, N. (1984), *Uneven Development*, Oxford: Basil Blackwell.
Stam, E. (2007), 'Why butterflies don't leave: the locational behaviour of entrepreneurial firms', *Economic Geography*, **83**, 27–50.
Storper, M. (1985), 'Oligopoly and the product cycle: essentialism in economic geography', *Economic Geography*, **61** (3), 260–82.
Storper, M. (1995), 'The resurgence of regional economies ten years later: the region as a nexus of untraded interdependencies', *European Urban and Regional Studies*, **2** (3), 191–221.
Storper, M. (1997), *The Regional World: Territorial Development in a Global Economy*, Cambridge, MA: Harvard University Press.
Storper M. and R. Walker (1989), *The Capitalist Imperative: Territory, Technology and Industrial Growth*, Oxford: Blackwell.
Sturgeon, T. (2000), 'How Silicon Valley came to be', in M. Kenney (ed.), *Understanding Silicon Valley: Anatomy of an Entrepreneurial Region*, Stanford, CA: Stanford University Press, pp. 15–47.
Sunley, P. (1992), 'Marshallian industrial districts: the case of the Lancashire cotton industry in the inter-war years', *Transactions of the Institute of British Geographers*, **17**, 306–20.
Sydow, J., G. Schreyogg and J. Koch (2005), 'Organizational paths: path dependency and beyond', Paper given at 21st EGOS Colloquium 30 June–2 July 2005, Berlin, Germany.
Taylor, M. (1986), 'The product cycle model: a critique', *Environment and Planning A*, **18** (6), 751–61.
Utterback, J.M. (1996), *Mastering the Dynamics of Innovation*, Cambridge, MA: Harvard Business School Press.
Veblen, T. (1898), 'Why is economics not an evolutionary science?', *Quarterly Review of Economics*, **12**, 373–397.
Vernon, R. (1966), 'International investment and international trade in the product cycle', *Quarterly Journal of Economics*, **80** (2), 190–207.
Walker, R. (2000), 'The geography of production', in E. Sheppard and T. Barnes (eds), *A Companion to Economic Geography*, Oxford: Blackwell, pp. 113–32.
Witt, U. (1998), 'Imagination and leadership: the neglected dimension of an evolutionary theory of the firm', *Journal of Economic Behaviour and Organization*, **35** (2), 161–77.
Witt, U. (2003), *The Evolving Economy*, Cheltenham, UK and Northampton, MA, USA: Edward Elgar.
Witt, U. (2006), 'Evolutionary concepts in economics and biology', *Journal of Evolutionary Economics*, **16**, 473–6.
Zook, M. (2005), *The Geography of the Internet Industry*, Oxford: Blackwell.

4 Complexity thinking and evolutionary economic geography

Ron Martin and Peter Sunley

1. Evolutionary economic geography: in search of conceptual foundations

Over the past few years the outline of a new evolutionary paradigm has begun to take shape in economic geography (see, for example, Bathelt and Boggs, 2003; Boschma and Frenken, 2003, 2006; Boschma and Lambooy, 1999; Essletzbichler and Rigby, 2004; Hassink, 2005; Lambooy and Boschma, 2001; Martin and Sunley, 2006; Rigby and Essletzbichler, 1997). Although this paradigm is still in its infancy, certain concepts and approaches have already assumed a prominent role in geographic-evolutionary interpretations of the economic landscape. In particular, most of the work towards the construction of an evolutionary economic geography has drawn on a particular version of evolutionary economics, which blends Nelson and Winter's (1984) evolutionary theory of the firm with neo-Darwinian (evolutionary biology) analogies and metaphors (especially variety, selection, novelty and inheritance).[1] Without doubt, the neo-Darwinian approach has been a highly influential perspective within evolutionary economics, and it is therefore not surprising that it has served as a major source of inspiration for evolutionary economic geographers.

But the neo-Darwinian framework is not the only possible one. Nor indeed, is evolutionary biology necessarily the most appropriate source of concepts, analogies and metaphors (Chattoe, 2006; Wimmer, 2006). In fact, there is debate within evolutionary economics itself about drawing on Darwinian concepts and the ideas of evolutionary biology.[2] And this is not a new debate. On the one side, for example, Alfred Marshall once famously concluded that 'the Mecca of the economist lies in economic biology rather than economic dynamics' (Marshall, 1930, p. xiv).[3] Others, however, have urged caution about borrowing ideas from biology. Penrose (1952, p. 819) for instance, argued that:

> In seeking the fundamental explanations of economic and social phenomena in human affairs the economist, and the social scientist in general, would be well advised to attack his [*sic*] problems directly and in their own terms rather than indirectly by imposing sweeping biological models on them.

More recently, even Hodgson (1993), a strong protagonist of an evolutionary approach to economics, has emphasised that while biological analogies are far more appropriate than the mechanistic ones that underpin mainstream economics, they have to be used cautiously. Further, as Lawson (2003) points out, there is the important question of whether it is ontologically meaningful to abduct notions from evolutionary biology into the socio-economic realm (see also Chattoe, 2006). In his view this question has yet to be adequately answered. Others, however, are more optimistic. For example, Witt (2003) advances a so-called ontological 'continuity hypothesis' of evolutionary economics, the suggestion that while the Darwinian notion of natural selection is only one form in which evolution occurs in nature, it is that form which 'historically, has shaped the ground and still defines the constraints for man-made, or cultural evolution' (Witt, 2003, p. 15). Somewhat between these two different viewpoints other authors have suggested that concepts such as variety, natural selection, inheritance and the like, do not have to carry over strict biological connotations when used as in economics, but can be used to identify 'generic' features of evolution that can be given specific meaningful economic interpretation (see Metcalfe, 1998; Witt, 1999). And so the discussion continues. All in all, it seems to us that the jury is still out on the question of whether a viable evolutionary economics – and thus by implication, a viable evolutionary economic geography – can be based solely on principles drawn from evolutionary biology.

It is not our intention to pursue this intriguing issue further here, however (for two such discussions, see Essletzbichler and Rigby, 2007; and Frenken and Boschma, 2007). Instead, our aim is to explore the potential scope – and limits – of a second approach to constructing an evolutionary perspective within economic geography, one based on what we shall call 'complexity thinking'. Interest in complexity and complex systems goes back at least to the 1940s, but in the 1970s and 1980s work on the dynamical properties and structural transformation of non-linear, 'far-from-equilibrium' systems in the natural and physical sciences led to the development of a new field that quickly became labelled as the 'science of complexity' or 'complexity theory' (Nicolis and Prigogine, 1977, 1989). Over the past two decades or so, this area of research has developed apace, focusing on, among other things, the evolutionary behaviour of 'self-organising systems', 'self-regenerating ('autopoietic') systems', 'complex adaptive systems', and 'complex evolutionary systems' (for example, Bak, 1996; Holland, 1992, 1995; Kauffman, 1995; Schweitzer, 1997). At the same time, 'complexity thinking' and 'complexity ideas' have diffused into several areas of the social sciences (Byrne, 1998), including not only economics (Anderson et al., 1988; Arthur et al., 1997; Metcalfe and Foster, 2004; Ramlogan and Metcalfe, 2006), but also economic and social history (Frenken and Nuvolari, 2004; McGlade, 2006), archaeology (Bentley, and Maschner, 2001, 2003), political theory (Rosenau, 1995), organisational and management theory (Stacey et al., 2000), and computing science (Bullock and Cliff, 2004).[4] It has had less impact on human geography, although recently it has begun to receive attention there too (for example, Gattrell, 2005; Harrison et al., 2006; Manson and O'Sullivan, 2006; Plummer and Sheppard, 2006; Thrift, 1999).[5] Such has been the growth of the field that some talk of a new episteme that challenges conventional epistemological and ontological assumptions about the nature and behaviour of natural and social phenomena (Wolfram, 2002).

This expanding interest and increased participation in the 'discourse of complexity' has not yet resulted in any clear, precise or generally agreed definition of the term, however; and to refer to complexity 'theory' is perhaps to exaggerate the degree of conceptual coherence and explanatory power associated with the notion. The main reason for this is that by its very nature as a holistic concept the notion of complexity resists easy reduction to a set of law-like statements or universal theoretical principles. As its two leading exponents put it, 'complexity is one of those ideas whose definition is an integral part of the problems that it raises' (Nicolis and Prigogine, 1989, p. 36). Nevertheless, there are some apostles of complexity who believe there are general principles that apply to all complex systems and that eventually it should therefore prove possible to construct a sort of unified theory of complexity. This has been the central thinking behind the Santa Fe Institute programme of research on the 'science of complexity'. The aim of the Santa Fe studies has been to explore the possibility of a formal theory of complexity – or more precisely, a theory of 'self-organisation' – that applies equally to both natural and social systems. Notable examples have included the work of Allen (1982, 1985, 1992) and Arthur (1988, 1989, 1994; and Arthur et al., 1997) on social and economic-technological systems.

Interestingly, Arthur's work in particular contained a number of applications to geographic phenomena, such as the development of urban and industrial location patterns. Subsequent to, and somewhat similarly to Arthur's work, Krugman has used complexity theory as one of the conceptual strands of his so-called 'new economic geography' (Krugman, 1994, 1996, 1997). Following the Santa Fe *Zeitgeist*, Krugman has sought to show how:

> Models of self-organisation can be applied to many economic phenomena – how the principle of 'order from instability' which explains the growth of hurricanes and embryos, can also explain the formation of cities and business cycles; how the principles of 'order from random growth' can explain the rules that describe the sizes of earthquakes, meteorites and metropolitan areas. (1996, p. vi)

The main argument underpinning Krugman's thesis is that common principles of self-organisation can be shown to operate across all sorts of systems – physical, biological and socio-economic – and that these principles provide a new view of how the economy structures itself in space and time.[6]

We should emphasise immediately where we agree and where we disagree with Krugman. A survey of the literature on complexity thinking does indeed suggest that natural, physical and social systems display certain similarities in 'complex behaviour', that is the emergence, under certain conditions, of self-organised complexity at a macroscopic scale in the form of spatial patterns or temporal rhythms. But the actual processes involved in the emergence of self-organised complexity obviously differ as between, say, cellular biology, the human brain, societal organisation and economic systems. This implies that distinct limits are likely to exist to the construction of a single, unified 'meta-theory' of complexity that is equally applicable to such diverse phenomena. Such a theory would perhaps only be possible at a very high level of abstraction and generalisation, which presumably is why some adherents of complexity thinking – including the Santa Fe school, and many others (such as Krugman, 1996) – seek to establish formal mathematical principles of complex behaviour.[7] However, it is our view that a formal (mathematical) modelling methodology is neither necessary nor of itself sufficient for understanding the complex behaviour of the economic landscape; evolutionary processes in the social-economic sphere are not easily reduced to, nor rarely can be adequately represented by, formal models.[8] Thus while we might share Krugman's view that the economic landscape can be viewed as a complex evolving system, we do not subscribe to the argument that this automatically requires the adoption of a model-based methodological strategy. Our task here, instead, is more ontological in purpose, namely to explore how far and in what ways some of the 'generic' aspects of evolutionary behaviour that are held to characterise complex systems can inform how we think about and conceptualise the economic landscape and its evolution. As in the case of borrowing ideas from evolutionary biology, so there are questions concerning the interpretation of metaphors and analogies transferred from 'complexity theory' to economic geography. The chapter is intended to throw some light on this issue. We begin, then, in the next section, with identifying some of these generic principles or properties that are held to characterise complex systems.

2. 'Complexity thinking': some generic concepts and principles

At the very outset, we need to distinguish between 'complexity' and 'complication'. The description 'complex' is often inappropriately attributed to systems by virtue of their having a large number of component parts, when these systems are merely complicated. A system of this kind can be understood by taking it apart and rebuilding it, like a clock or a car – the system is explicable through a description of is component parts. On the other hand a system is complex when it comprises non-linear interactions between its parts, such that an understanding of the system is not possible through a simple reduction to its component elements. A complicated system, then, need not be complex, in the sense of exhibiting complex behaviour: of course, a complex system may also be complicated.

As mentioned above, while there is as yet no generally agreed set of well-defined 'law-like' statements that together constitute a universal theory of complexity, nevertheless what distinguishes complex systems is the way they exhibit emergent self-organising behaviour, driven by co-evolutionary interactions, and an adaptive capacity that enables them to rearrange their internal structure spontaneously (Pavard and Dugdale, 2000). More specifically, seven generic properties can be identified as characteristic of complex systems (see Table 4.1).

First, a complex system has a *distributed* nature and representation, in the connectionist sense, whereby the system's resources are physically or virtually distributed across various sites, and its functions and the relationships and feedbacks that exist among its elements occur over various spatial ranges and scales: complex systems are characteristically multi-scalar. Second, it is often supposed that a system changes only inside its own frontier: this is the definitional notion of 'operational closure'. In contrast, it is typically difficult to determine the boundaries of a complex system: the boundary between a complex system and its environment is usually dependent on the purpose of the analysis, or on the context, and not on any intrinsic property of the system itself. In short, *openness* (or *non-isolation*) is an inherent feature of complex systems. This in turn is closely related to the idea that such systems tend also to be *dissipative*, in the sense that they are in constant interaction and exchange with their environments, and thus experience a continual inflow and outflow of energy, matter and information. When these characteristics are combined with *non-linear dynamics*, arising from the mutually reinforcing feedbacks among a complex system's parts, the result is an *irreversibility* of change and a tendency towards path dependence in the system's trajectory and behaviour. At the same time, however, openness implies a susceptibility to externally induced fluctuation and perturbation, and such forces can cause a shift to a new regime. In other words, 'the passage to complexity is intimately related to the *bifurcation* of new branches of solutions [trajectories] following from the *instability* of a reference state, caused by the nonlinearities and [environmental] constraints acting on an open system' (Nicolis and Prigogine, 1989, p. 73; original emphasis). Because of its inherent connectivity, non-linearity and

Table 4.1 Some key generic properties of complex systems

Property	Attributes
Distributed nature and representation	The functions and relationships are distributed across system components at a whole variety of scales, giving the system a high degree of distributed connectivity.
Openness	The boundary between a complex system and its environment is neither fixed nor easy to identify, making operational closure dependent on context (and observer). Such non-isolated systems tend to be dissipative – subject to constant interaction and exchange with their environments.
Non-linear dynamics	Complex systems display non-linear dynamics because of various complex feedbacks and mutually self-reinforcing interactions amongst components. Complex systems are thus often characterised by path dependence.
Limited functional decomposability	Because of its high degree of connectivity, and the open, dynamic nature of its structure, there is limited scope for decomposing a complex system into stable components
Emergence and self-organisation	There is a tendency for macro-scale structures (including spatial structures) and dynamics to emerge spontaneously out of the micro-scale behaviours and interactions of system components.
Adaptive behaviour and adaptation	The same processes of self-organisation imbue complex systems with the potential to adapt their structures and dynamics, whether in response to changes in the external environment, or from within through co-evolutionary mechanisms or in response to 'self-organised criticality'.
Non-determinism and non-tractability	Complex systems are fundamentally non-deterministic. It is not possible to anticipate precisely their behaviour even if we completely know the function of their components. This does not imply, however, that the behaviour of such systems is random, in the sense of being haphazard.

Sources: Based on Horgan (1995), Pavard and Dugdale (2000), Adami (2002), McGlade and Garnsey (2006), among others.

openness, a complex system also affords *limited functional decomposability*; that is, its overall (macro-level) functioning cannot be deduced from knowledge of the function of its sub-components.

Yet further, the pattern and nature of interactions (non-linear and distributed) between the elements of the system allow it to functionally restructure itself over time. Indeed, the properties of *self-organisation and emergence* are often held to be the primary distinguishing features of complex systems.[9] These ideas refer to the observation that there are many systems that at one level can be described as consisting of numerous micro-scale components that are individually 'simple' and that interact with each other in limited, simple ways, often only influencing neighbouring components, yet at another scale are able to exhibit some complex overall system-level structure and behaviour. In broad terms, it is those systems that exhibit the 'emergence' of macro-scale structure and dynamics spontaneously from the micro-level behaviours and interactions of their components that are referred to as 'complex systems'. These same processes, including co-evolutionary mechanisms among different system components and levels, give complex systems an *adaptive quality*, whereby their structures change in response either to changes in a system's external environment, or from within through what some have termed 'self-organised criticality' (Bak, 1996), in which the system evolves to a particular 'critical' state that then generates chain reactions between components to produce a major change in the system's structure and/or dynamics (a sort of 'punctuated equilibrium' form of system evolution).

Finally, because of these various attributes, complex systems are fundamentally *non-deterministic*. Even if we have complete information on the function and interrelationships of components, it is not possible to anticipate their behaviour precisely. Complex systems are inherently stochastic in nature. But that does not mean their behaviour is random in the sense of being haphazard. There are causal processes at work, but they operate through complex, distributed feedback and self-reinforcing mechanisms that are unlikely to be detected by standard measures of association (correlation) between assumed determinants and presumed effects (McGlade and Garnsey, 2006).

According to some exponents, together these various characteristics form a sort of 'vocabulary of complexity' (Lissack, 1999; Nicolis and Prigogine, 1989). In effect, they constitute a 'complexity ontology', a particular view of how reality (from physical systems to biological systems to social systems) is structured and behaves (evolves). Of itself this ontology does not imply a specific set of methods or 'tools' of analysis, though as mentioned above, many advocates of complexity do seek to express complex systems and their dynamics in terms of formal mathematical models. This distinction between the ontological and methodological aspects of complexity thinking is an important one, as developments and debates within economics demonstrate.

3. Economics and complexity thinking

In recent years complexity theory has attracted increasing interest from economists, and is now believed by many to be a novel and powerful framework of thought capable of challenging the fundamental principles of the mainstream economic canon (see, for example, Arthur et al., 1997; Auyang, 1998; Colander, 2000a, 2000b; Krugman, 1996; Metcalfe and Foster, 2004; Potts, 2000; Rosser, 2004; Schenck, 2003).[10] According to one of its main advocates, 'complexity changes everything; well maybe not everything, but it does change quite a bit in economics' (Colander, 2000a, p. 31). It is suggested that complexity ideas 'are beginning to map out a radical and long-overdue revision of economic theory' (Buchanan, 2004, p. 35). In one such manifesto, for example, Beinhocker (2006) suggests the term 'complexity economics' as an umbrella for a number of streams of theoretical and empirical work that can be linked directly or indirectly with 'complexity thinking'. While he emphasises that 'complexity economics' is still more a research programme than a single synthesised theory, he identifies five key dimensions – or what he calls 'big ideas' – that mark out 'complexity economics' from 'traditional economics' (Table 4.2): the economy as an open, non-linear system; made up of agents with bounded rationality, who learn and adapt; who interact through constantly changing networks; whose micro-behaviours and interactions are the source of emergent pattern and order at the macro-level; and who are the source of the constant novelty that imbues the economy with its evolutionary momentum.

Table 4.2 Beinhocker's five 'big ideas' that distinguish 'complexity economics' from traditional economics

	Complexity economics	Traditional economics
Dynamics	Open, dynamic, non-linear systems, far from equilibrium	Closed, static linear systems in equilibrium
Agents	Modelled individually; agents use inductive rules of thumb to make decisions; have incomplete information; are subject to errors and biases; and learn and adapt over time	Modelled collectively; agents use complex deductive calculations to make decisions; have complex information; make no errors and have no biases; and have no need for learning and adaptation (are already perfect)
Networks	Explicitly models interactions between individual agents; networks of relationships change over time	Assumes agents only interact indirectly though market mechanisms
Emergence	No distinction between micro- and macro-economics; macro patterns are emergent result of micro-level behaviours and interactions	Micro- and macro-economics remain separate disciplines
Evolution	The evolutionary process of differentiation, selection and amplification provides the system with novelty and is responsible for growth in order and complexity	No mechanism for endogenously creating novelty or growth in order and complexity

Source: Beinhocker (2006, p. 97).

The problem, however, is that different economists approach the idea of 'complexity' in quite different ways. According to Perona (2004), two sharply divergent versions of 'complexity economics' can be distinguished: 'theoretic' and 'ontic'.[11] The term 'theoretic' refers to those versions of complexity economics that treat non-linearity, instability, self-organisation, emergence, adaptive behaviour and the like, as features of the mathematical *model*, the equation or system of equations, used to *represent* a particular type of reality. Perona contrasts this to 'ontic' versions of complexity economics, that is where complexity is seen as a property of real economic systems and phenomena. Though we agree with the general thrust of Perona's argument, to our mind the label 'theoretic' in this context is not altogether helpful, since not all theories are cast in terms of mathematical deductive models; and, some might argue, what is an ontology if not a particular theory of how the world is structured? What Lawson (2003) elsewhere refers to as a distinction between 'scientific-ontological' and 'social-ontological' approaches to complexity might be more appropriate, where the former is concerned with the nature or structure of entities posited or presupposed by some scientific model, and the latter is concerned with the nature of socio-economic existence apart from any such model. Arguably, the key distinction is between epistemological (methodological) and ontological approaches to complexity. In complexity theory, epistemology has dominated over ontology, with the latter often reduced to the former (in the manner discussed by Lawson). The assumption has been that the same formal (thermodynamic and chemico-physical) models can be applied across numerous different fields, including the economic.

In Perona's view, the efforts of Arthur (and, we would add, Krugman) to construct a complexity economics are of this model-based ('theoretic', or 'scientific-ontological') kind. Thus we find both authors postulating an abstract economic landscape that is viewed as an adaptive system, consisting of many agents who interact continuously among themselves, while being allocated to a pre-given geometry of locations and immersed in an overall 'evolutionary' environment (Arthur, 1994; Arthur et al., 1997; Krugman, 1994, 1996), all represented by some system of non-linear equations, power law functions and the like. In these models, agents' interactions and location decisions are usually described by means of (predetermined non-linear) rules, which are repeated through several steps, thus accounting for the system's 'dynamic' nature and spatial development. Complexity in these treatments is not a characteristic of how causal processes are connected to each other, but a characteristic of the behaviour of the spatial-temporal series of a particular variable, typically represented not by actual data but by the solution sequence of a computational or simulation model (see not only Arthur, 1994; Arthur et al., 1997; and Krugman, 1997; but also Rosser, 2004). According to Viskovatoff (2000), although this approach to complexity economics (and complexity 'economic geography' as formulated by Arthur, Krugman and other economists) may well abandon some of the artificial assumptions of neoclassical economics (such as the rational behaviour of agents, and an inherent tendency to a single, unique equilibrium), it remains committed to deductive theorising, in this case as formulated by non-linear models (and, it should be added, to the idea of equilibrium, even if the latter is now no longer unique but dependent on the 'initial conditions' or 'starting point' of the economic system). In effect, the complex economy becomes identified with the simulation model and solution sequence used to represent it.

The problem is that there are far fewer examples of an ontological ('ontic') perspective. One of the most concerted efforts in this direction is that of Potts (2000), for whom complex systems theory is the most suitable basis for constructing an evolutionary economics:

> The hypothesis of evolution towards complexity is a conjecture to the effect that a balance between order and chaos, between stasis and change, is the ultimate principle underlying all evolutionary processes. Where equilibrium is the expression of 'balance' in an inert, mechanical world of point-like existence, complexity is the expression, the structural signature, of balance in a world of interacting dynamic systems. The hypothesis of evolution towards complexity is the logical principle that interlinks the geometry of all economic systems. (Potts, 2000, p. 91)

Following Kauffman's (1993) contention that there is no strong reason to attribute the emergence of order in biological systems solely to the force of selection, Potts argues that much of the order and coordination in an economic system may not be the result of 'market selection' at all, but a spontaneous order of self-organised systems. According to Potts, the concept of complexity – and more specifically the hypothesis of evolution towards complexity – contains within its meaning a number of high-level connecting principles that are prominent in heterodox economic theories: 'evolutionary economics is an eclectic rubric centred around the paradigm of the complexity of open systems processes, and its basic substance is both more encompassing and more protean than a simple transferral of metaphor' (2000, p. 186).

For Potts, the ontology of methodological individualism, social atomism and equilibrium of traditional economics is replaced in complexity economics by an ontology of 'connectivity':

> Generally connections are specific direct relationships between elements and are ubiquitous in the economic system. They exist in the structure of interdependencies and interactions between agents. They exist in the modalities of technology and the forms of organisation and competence. They exist as contracts. They exist in the structure of decision rules and the way that information is processed, In all such events, the dynamics of economic systems can be seen to occur most in the space of connections. (Potts, 2000, p. 3)

Potts goes further and suggests that knowledge in all of its multifarious forms is in fact about connection. For Potts, as for many evolutionary economists, it is the growth of knowledge that drives the process of economic evolution (Foster, 1993, 1994, 1997; Metcalfe, 1997, 1998; Metcalfe and Foster, 2004). Similarly, Foster pieces together a vision of evolutionary economics based not on biological analogy, but on principles of complex self-organisation in which knowledge plays a formative role: 'economic self-organisation involves acquired energy and acquired knowledge, which in combination, yield creativity in economic evolution' (Foster, 1997, p. 444). According to such authors, the structure of knowledge is a structure of connections, and the various instances of knowledge – such as technology, routines, habits, competences and the like – are instances of specific connections, that is of networks, that seem to work in a particular environment. Thus, from this perspective, economic evolution – and hence complexity economics – is about the emergence and evolution of multiple connections in the form of networks.[12]

However, notwithstanding the efforts of Potts, Foster and others to forge a 'heterodox' complexity economics, for the most part much of complexity thinking in economics commits the conflation referred to by Lawson, that is it reduces the ontological to the epistemological. The economy is taken as constituted by scale-invariant laws or dynamics that apply not just to the economy but to all sorts of complex systems: self-organising systems reach a state of spontaneous order and then 'follow a dynamic pattern that is lawful in its own right' (Batten, 2001, p. 94). Typically these laws and dynamics are identified by the presence of behavioural and statistical signatures or hallmarks such as power laws, punctuated equilibria, and phase transitions, and attractors (see Krugman, 1994).[13] In many cases, it is the belief in these universal complex dynamics that is used to justify both mathematical formalism and, in an economic context, agent-based computer simulations.

The belief that different types of complex system all show signature behaviours and common emergent patterns runs through all versions of complexity economics to some degree. Yet, such a conceptualisation overlooks the obvious but important fact that social and economic systems show different and distinctive forms of complexity: social, spatial and historical context shape the nature of the economy and its behaviour. According to Foster (2005), all economic systems are 'third-' or 'fourth-order' complex.[14] In 'third-order' systems individual agents modify their environments to increase access to energy and resources, reduce their effort costs, or seek out free information to increase their knowledge. In 'fourth-order' complex systems individual agents form beliefs about the beliefs of others (collective understandings) that are critical for their survival.

More specifically, in the modern economy they seek out information about the beliefs/knowledge of others in order to cooperate to modify the environment. This implies that there will particular forms of complexity in economic systems that are beyond any 'natural laws' of complexity but are highly social, institutional and historical. However, this is only a starting point. Foster goes on to argue that 'Economic self-organisation is not the same as biological self-organisation, despite the fact that they share common properties' (2000, p. 325). The principles and rules of first and second order complex systems can be applied to economic systems, but they operate 'somewhat differently' because of the greater role of knowledge and foresight (Foster, 2004, p. 11).

In similar fashion, complexity economics also synthesises evolutionary approaches with complexity thinking on dissipative systems. In Beinhocker's (2006) account, for example, the essential complexity framework that spans both natural and social systems is an adaptive, iterative evolutionary algorithm. In this algorithm, the universal evolutionary mechanisms are those that act to differentiate, select and replicate, and allow adaptive agents to co-evolve with their environments. He suggests that there has been 'too much loose analogizing about how the economy might be *like* an evolutionary system' (p. 12). He adds:

> Modern efforts to understand the economy as an evolutionary system avoid such metaphors and instead focus on understanding how the universal algorithm of evolution is literally and specifically implemented in the information-processing substrate of human economic activity. While both biological and economic systems share the core algorithm of evolution and thus have some similarities, the realizations of evolution are in fact very different and must be understood in their individual contexts. (Beinhocker, 2006, p. 12)

Thus in recent complexity economics, most authors have been at pains to emphasise that economic systems display forms of complex behaviours that are distinct from, and yet build on and resemble, the versions of complexity identified in the physical and biological worlds. But the argument that there are both shared common patterns and yet very distinctive social and economic versions of evolutionary complexity, creates several tensions and unresolved questions. Complexity economics insists that economic systems are distinctive, yet, as we will see, when convenient, it continues to import complexity dynamics and models identified in natural science and suggests that analogous patterns can be identified in the economic system. How much complexity science can we translate into the economic sphere, and how exactly should it be mediated and modified? How many of the principles and axioms that are distinctive and characteristic of complex natural evolutionary systems carry over and extend to social forms of complexity? Given the lack of agreement on such questions, it is not surprising – as Perona points out – that while complexity ideas are thriving in economics, there remains confusion between epistemological and ontological aspects of complexity economics. A complexity approach to evolutionary economic geography will need to address and resolve this confusion if it is to provide a convincing new approach. In addition, as we now go on to discuss, when we move from the economy to the economic landscape there are further issues that need clarification.

4. Emergent economic landscapes?

Complexity economics at present says very little about space and geographical differences. Admittedly, existing models of self-organising economic landscapes demonstrate that relatively simple rules of agglomeration and centrifugal forces can produce emergent spatial orders that resemble edge cities and clusters or show statistical regularities (Arthur, 1994; Batty et al., 2004; Krugman, 1994, 1996). At best, they may generate selected 'stylised facts' in economic geography and 'space' appears in these models as a patterned outcome, but the models tell us little that is new about complex economic change and system adaptation over space. And the nature of geographic space in these modelling exercises is either wholly abstract or empirically crude. As noted earlier, some of the characteristics of non-linear, path-dependent system change have also been applied to geographical entities such as clusters, regions and nation-states (see Arthur, 1994; Garnsey, 1998; Krugman, 1996; Martin and Sunley, 2006). However, we should start by carefully considering whether such spatial entities can legitimately be described as complex adaptive, self-organising systems.

The dynamics of complex adaptive systems depend on their configuration and how this responds to innovation and shocks. Such systems are constituted by interconnections across multiple scales between diverse and heterogeneous agents. Many of them display spatial behaviour that bears crucially on their robustness and adaptability. Moreover, local co-evolutionary effects within such systems typically produce innovation and novelty (Markose, 2005; Metcalfe et al., 2006). Indeed, it has recently been argued that complex systems are preternaturally 'spatial' (Thrift, 1999). O'Sullivan et al. (2006), for example, argue that spatial variability is central to complexity, as where elements are located relative to other elements is critical to their behaviour (Manson and O'Sullivan, 2006). As Bullock and Cliff (2004) have argued, however, the role that spatiality plays in underpinning complex adaptive behaviour is poorly understood. While many of the leading accounts of complexity and complexity economics discuss system movements in 'state-spaces' and their adaptive walks on 'fitness landscapes', they say little about geographic space and its relation to the adaptive behaviour of individuals and businesses.

If there is a spatial metaphor essential to complexity economics then it is the 'network'. Complexity approaches represent 'the economy' as made up of innumerable flows and connections, and they are predicated on the notion of incomplete and selective networks. The starting point is that everything is not connected to everything else so that the 'force field' metaphor underlying neoclassical economics is inappropriate (Potts, 2000). Instead, the bounded rationality and imperfect knowledge of economic agents mean that we should address imperfect and incomplete networks that are irreducibly broken and partial, and in this sense they are spatially distributed, with a bias towards some degree of localisation. Complexity thinking tends to start with the assumption that components interact most strongly with their nearest neighbours, and in some physical systems this means a form of distance decay. The idea of complex territorial economies is then most aptly applied where interactions between agents are geographically localised. But, of course, where distance is as much or more 'relational' rather than spatial, 'localised' interactions do not always translate into spatial proximity. As Cilliers (2005a, 2005b) has argued, we can not always assume that socio-economic subsystems are spatially contiguous. Parts of socio-economic systems may exist in totally different spatial locations and systems may interpenetrate each other and so be part of different systems simultaneously. Localised interactions can then simply mean interaction with selected other components. Ultimately, it is plausible to argue that spatial entities such as regions and cities become self-organising complex systems when they are strongly interactive and the interdependencies between agents are 'global' in the sense of operating throughout the trading system (Batten, 2001). For example, Batten (2001) argues that medieval Europe in the Carolingian era was a weakly interactive system in which villages and groups of villages were autarkic and self-contained. However, as long-distance trade and the geographical circulation of goods and merchants intensified, this weakly interactive economy was transformed into a highly interactive one, and the European economy self-organised. Does this mean, however, that all modern economies are self-organising? What are the critical transitions and thresholds in different types of interaction and circulation that define whether a region or city is strongly rather than weakly interactive?

A defining feature of a complex system is that it is composed of interacting sub-systems and hierarchical levels, and on this basis we might argue that the national economy can be divided into smaller territorial sub-systems such as regions, cities and localities. Batty et al. (2004), for example, claim that cities are the quintessential example of complexity. But if complex systems define their own scales and can shift scales unpredictably, then can we assume a certain geographical scale can always be defined as a complex system? The boundaries of complex systems are never pre-given but are a matter of framing, which is partly about the purpose of study and research strategy, and partly about the constraints arising from the operations of the system itself (Cilliers, 2005a, 2005b). Dividing a complex system into territorial sub-systems is bound to be at the cost of analytically fragmenting and simplifying its complexity, but it can be justified if the constituent flows and connections are producing some forms of identifiable boundary and system integrity. Clearly, identifying spatial economic sub-systems does not imply that they are closed. Rather the boundaries of complex systems are neither fully open nor fully closed. Moreover, Cilliers (2001) also argues that when components are richly interconnected, there is only a short route to the outside of the system. There is no safe inside; the system boundary is folded in so that in a richly interconnected system we are never far away from the boundary. Boundaries are functional constitutive components that do not separate but connect systems with their environment. Sub-systems can be identified within complex systems if the sub-system becomes responsible for maintaining its own configuration, and it can then be said to show 'organisational closure'. In this state the organisation is determined purely internally even though the sub-system (a region or city, say) exchanges energy and matter (that is, flows of goods, services, knowledge, capital, money and people) with its environment (other regions and cities). The interactions between these sub-systems (regions, cities, clusters, etc.) may also determine systems at a higher hierarchical (and spatial) level so that a 'boxes within boxes' architecture emerges.

The subsequent question is whether we can justifiably argue that spatial entities such as regional economies and clusters are economic sub-systems that show emergent properties. Do they emerge from unplanned and unintentional localised interactions between agents and display relatively ordered paths of change? What is not clear here is whether such units (clusters, cities, regions) show 'organisational closure' such that they retain their organisational identities and configuration despite fluctuations and perturbations in their external environments. Is it feasible that regions, clusters and cities develop key internal interactions that provide them with some organisational 'attractor' and continuity in the face of environmental changes? Given that market-based economic systems are always importing knowledge and information then there is no simple distinction between internal organisation and flows, and the metaphor of organisational closure may imply too great a level of coherence, stability and endogeneity. Furthermore, what has not been explained so far is how we might identify such organisational maintenance in regional and local economies. What exactly are the key organisational connections, flows and interactions that define this condition and are they flows of money, value, people or ideas?[15] A further question is whether we can apply the idea of emergent properties to economic places. According to O'Sullivan et al. (2006) 'the novelty of complexity lies in a sustained attempt to grapple with the bottom up emergence of aggregate behaviour on the one hand, and the top down impact of emergent structures on the behaviour of constituent elements on the other' (p. 614). In this view, emergent characteristics exercise 'downward causation' and influence the localised interactions between agents (Gilbert, 2002, refers to this as 'second-order' emergence). There is now, of course, much evidence that there are forms of positive feedback in regional and cluster development so that initial advantages and firm location decisions may initiate path dependent, self-reinforcing trajectories, and indeed Krugman (1996) argues that such feedbacks and spillovers are good examples of emergent effects.

However, there are also difficulties in applying the notion of emergence to economic landscapes, largely because of the key role of knowledge. One of the conditions of emergence is that local agents are unaware of the emergent effects of their actions and relations (Manson and O'Sullivan, 2006). Complex emergent effects are generated by relatively simple behavioural rules and interactions between agents. Clearly this does not apply in the same way in Foster's (2005) 'fourth-order' complex systems where action is to a far greater degree reflexive, intentional and subject to continuous monitoring and cooperative behaviour. As Sawyer argues, because of the character of human symbolic communication, processes of emergence in social systems are different from those in natural and biological systems: 'In social systems the components (individuals) contain representations of the emergent macropatterns, unlike in any other complex system' (Sawyer, 2005, p. 26). Thus, for example, to describe industrial clusters as emergent phenomena may obscure some of the ways in which the knowledge and expectations that maintain their success are not simply localised interactions but are also public, collective and interpretive. Economic knowledge may itself be continuously emergent, arising from exchanges of information (Potts, 2000; Ramlogan and Metcalfe, 2006). But if that knowledge then acts as the basis for further change and economic growth, then the idea of simple sequences of localised interactions generating emergent effects would seem too simplistic.

5. Self-organisation in economic landscapes

As we have noted above, a key set of issues in complexity economics revolves around the notion of self-organisation. While some complexity economists use the notions of self-organisation and self-transformation fairly loosely, others hardly use the concepts at all (see, for example, Beinhocker, 2006). Recall that self-organisation describes the way in which systems order themselves without central direction or external control such that they acquire and maintain structure and arrange selected parts to promote (and reproduce) a specific function. In this latter respect, the idea of self-organisation is closely related to the notion of autopoiesis, which refers to the dynamics of a non-equilibrium system that produces the components that in turn continue to maintain the organised structure that gives rise to these components. At one level, these twin concepts would seem highly applicable to the economic landscape. Cities or clusters, for example, would appear to be good examples of self-organised, autopoietic systems: as localised economies, they comprise sets of components (firms, institutions, infrastructures, people, etc.) that generate outcomes (decisions, relations, daily behaviours, profits, incomes, knowledge, and the like) that serve to reproduce the components that make up the city or cluster. The emergence of positive localised externalities or agglomeration economies might thus be interpreted in autopoietic terms, since they feed back to help reproduce and maintain the economic components (firms, workers, institutions) whose spatial juxtaposition and inter-relationships create those very externalities. In this way, as organised (and dissipative) autopoietic systems, cities, clusters and particular types of regional economy can remain stable for long periods of time despite people, goods, knowledge, and money continually flowing into and out of them.

Yet there are also clearly problems using a self-organisation metaphor in economic geography. First, there is the question of how valid it is to think of the economy of a city, or cluster, or region in autopoietic terms. City and regional economies are not internally coherent structures: certain components can be added or removed without necessarily influencing the organisational stability of the city or region as a whole; many firms may have few if any links with other local firms; and different parts of the city's or region's economy may function in different ways, and be linked to the external environment (external markets) in different ways. In other words, self-organisation in the economic landscape may not necessarily of itself equate with autopoietic dynamics.

Second, the basic assumption of distributed and dispersed control among system components is clearly not applicable to many types of economic organisation such as cities or regions. In fact, complexity economics says little about the power inequalities that exist in all economic landscapes and strongly shape the selection of institutional and organisational configurations. This has major implications, of course, as the assumption that the connections and configurations that exist in economies have been selected for their 'fitness' by market processes and for their ability to maximise flows of value, can yield a remarkably uncritical view. In fact connections and linkages in the economy are likely to be selected according to several different criteria simultaneously, including the vested interests of more powerful groups and their ability to channel and control these flows.

Third, self-organisation is problematic as it is hard to identify mainly endogenous dynamics when the boundaries of economic systems – such as regions and cities – are so hard to identify.[16] Metcalfe, for example, argues that the economy self-transforms when economic agents become dissatisfied or concerned about their returns so that they search for new ways of doing things and combine knowledge in new ways to produce new value flows. But it makes little sense to insist that these new bits of knowledge and problem solutions are mainly internal to the economy as human learning is far too tangled and unbounded for this. Furthermore, while markets may show forms of self-regulation and coordination, it is clear that there are numerous institutional and political preconditions which allow these coordinating effects to occur.[17]

Fourth, there are further questions about the relationships between self-organisation, connectivity and order in the economic landscape. Complexity theorists see self-organisation as a critical balance between order and chaos, and, according to Potts (2000), in the economic sphere the degree of connectedness is key to understanding the nature of this balance. An ordered system is defined as one with a low level of connectivity. In such a system a change at any one point or in any one component has limited impact on the rest of the system, which, as a whole, remains virtually unchanged: low connectedness implies a high degree of order and a high degree of stability of the system. In a system with a high degree of connectedness, a change at any point in the system impacts on many other elements so that change is propagated across the system as a whole. In the extreme case, where every element is connected with every other, the system would be wholly unstable, in effect chaotic. Complex systems are somewhere in the middle: 'order is associated with low connectivity, chaos with high connectivity, and complexity forms a narrow window of low-to-intermediate connectivity between order and chaos' (Potts, 2000, p. 90). According to Potts, the economy, as a complex system, is in dynamic balance between order and chaos, and is:

> Neither ordered not chaotic but both, in a balance between information, pattern and coordination being usefully locked into a system (as preferences, technology, institutions) and the continued experimentation and search for new patterns and the maintenance of flexibility within the system so that these may then be adapted. (2000, p. 90)

What does this imply for the economic landscape, for the geographies of the economy? Most economic geographers would agree that there is a high degree of stability and order to the spatial structure of the economy: cities do not appear or disappear overnight; patterns of industrial location and agglomeration do not change from one day to the next; regional specialisations do not develop or decline spontaneously. Yet this stability and order is hardly the result of low levels of connectivity, either within or between cities and regions. Indeed, adherents of the new 'relational turn' in economic geography would probably argue that it is primarily in terms of multifarious relational (connectivity) networks that the geographies of the economy should be understood. Yet again, equating order, stability and lock-in with low levels of connectivity would seem to run counter to those accounts that explain regional lock-in in terms of excessive inter-relatedness – the 'weakness of strong ties' argument (for an extended discussion of regional lock-in, see Martin and Sunley, 2006).[18] In short, the relationships between self-organisation, connectivity and order would seem to require extensive elaboration in an economic-geographic context.

And, fifth, so too does the relationship between self-organisation and adaptation (Essletzbichler and Rigby, 2007). According to Hodgson and Knudsen (2006a) self-organisation alone cannot provide the basis for an evolutionary economic theory as it tells us nothing about selection: 'Self-organization alone cannot explain the adaptation and differential survival of self-organized systems' (2006a, p. 16). In their view an exclusive focus on self-organisation at the expense of selection does not explain how systems are adapted to their environments. They conclude therefore that selection acts on self-organised systems once they have emerged. But whether the emergence of connections and interlinkages in economic systems (self-organisation) can be clearly distinguished from their subsequent selection is debatable. The contexts for the generation of variety and its selection in economic organizations are often combined and inseparable (Loasby, 2001). As we have already seen, other complexity economists appear to suggest that some complex systems have properties that make them robust and adaptable. They argue that self-organised systems move to a poised state, called the 'edge of chaos', where they balance between no order and too much order (Potts, 2000). In this critical state, complex systems have a dynamic efficiency that allows them to meet evolutionary and competitive pressures, whereas excessively ordered as well as chaotic, unstable systems are both eliminated. Self-organised systems have a high degree of resilience and robustness as they are marked by distributed and dispersed, rather than centralised, control, as well as by strong positive and negative feedback loops, and by a high level of redundant variety (Heylighen, 1999). It is not surprising then that this vision of complex systems has been used in a normative fashion and has been used to explain the resilience and adaptability of complex industrial clusters (Lindsay, 2005).

At the same time, however, complexity economics also suggests that once co-evolutionary complexity, in terms of a system's internal and external co-evolutionary linkages, passes a certain threshold, the system may become unresponsive to environmental pressures. Foster (1997), for example, points out that all dissipative systems have a tendency to degrade through time and the renewal of their links depends on the continual import of information, energy and resources. He also argues that complex systems tend to become more specialised as their order, integration and 'knowledge' increase. As their coherence increases, they specialise in adapting to particular environmental niches and if these niches suffer from resource depletion or the entry of new competitors, then even complex systems may disintegrate and start the process of self-organisation anew. Thus complex linkages and connections that constitute complex economic systems may over time prove to be too specialised. In the case of 'complexity catastrophes', too many interdependencies act to trap a system within a basin in a fitness landscape so that environmental selection can not operate. McKelvey (1999) applies this thinking to firms and Beinhocker (2006) also argues that hierarchical systems in organisations are often more efficient and adaptable because they can make and implement decisions more rapidly.[19] Once again however, whether this can be applied to the evolution of urban and regional economies is unclear. How can we judge when their connectivity has exceeded a beneficial value and started to move towards a complexity catastrophe? What would an excess of external connectivity mean in this case? Under this logic also, the increasing complexity of many production networks and regional economies might imply that we will witness more 'complexity catastrophes' in future, although this will depend, of course, on whether any such effect is offset by other trends in innovation and the evolution of knowledge.

6. Complexity and regional economic evolution

An issue of particular interest in the present context, given the above discussion and our comments in the introduction, is whether and how far complexity economics represents an advance on other types of evolutionary economics, and evolutionary economic geography, that draw insights by using Darwinian and other natural evolutionary analogies and metaphors. While some complexity economists argue that their approach is macroscopic and supersedes a microscopic, generalised Darwinianism (Foster, 1997), as we have already noted, most argue that complexity economics can synthesise complexity with evolutionary approaches based on natural selection. At the same time complexity economics tends to claim that it no longer has to rely on restrictive natural analogies and metaphors. Instead it aims to tell stories about its own subject matter (Wakeley, 2002). Beinhocker (2006), for example, argues that his approach is based on a universal evolutionary algorithm so that he does not have to rely on natural analogies. However, natural metaphors continue to be permeate complexity economics and are used when convenient. Beinhocker (2006) later identifies general laws of evolutionary systems that subsume the biological:

> The claim of the modern algorithmic view of evolution is that evolutionary systems are a universal class with universal laws. We can then ask whether the economy is part of that class and subject to those laws. If the answer is yes, then the economic and biological worlds are both members of that universal class. They may be very different in their implementations of the algorithm, and thus asking what a parent and an offspring are in economics makes no sense. Nonetheless, the two worlds are still subject to the same general laws of evolutionary systems, thus explaining the strong (pardon the metaphor) family resemblance. (p. 217)

Indeed, he relies heavily on finding a business equivalent to DNA which can be differentiated, selected and replicated. But do we need to search for this if socio-economic evolution is fundamentally different from natural evolution? Moreover, he also uses evolutionary psychology to explain what he means by wealth as 'fit order'. Goods and products that push people's 'pleasure buttons' are those that satisfy the needs and instincts derived from our long evolutionary history. Thus, his explanation proves reliant on drawing out the implications of natural evolution.

Furthermore, other complexity economists are more willing to use direct natural evolutionary analogies. For instance, one of the most interesting aspects of self-organisation identified in the complexity literature is the 'Red Queen' (or competitive co-evolution) effect in which two competing species become locked into an intensive and adaptive race, equivalent to running in order to stand still. There have been several applications of this idea in complexity economics (for example, Markose, 2005; Robson, 2005). Thus it has been argued that this effect may explain the dynamics of innovation in high-technology and financial sectors as it forces competitors to continuously introduce new variety. It is not hard to envisage a spatially defined version of Red Queen effects in which firms in clusters are driven to innovate more by the pressure of close competitors (Porter, 1998). The synthesis of evolutionary ideas with complexity means that recent versions of complexity economics still appear to be vulnerable to criticisms of evolutionary approaches. One of these is that they still portray human agents as mainly adapting to their environments rather than actively making these environments. There seems to be little in complexity economics on the ways in which powerful firms transform their economic environments through buying out competitors, switching investments into new sectors and locations, introducing major innovations in processes or products, or remaking markets through the effects of mass advertising.

Notwithstanding this continued use of the natural metaphor, the main difference between complexity economics and neo-Darwinian views of economic evolution appears to rest on the relative importance of system self-transformation relative to selection.[20] The key point here is that some complexity accounts argue that selection cannot operate on complex economies. This is either because there is too much innovation and mutation happening simultaneously or because interdependencies between system components mean that the fitness landscape is flat and that these co-evolutionary pockets become trapped in suboptimal fitness (Kauffman, 1993; McKelvey, 1999). In these conditions environmental pressures will not work even if they are strong. This surely implies that our analysis of the evolution of an economic landscape should look at both types of change and should not always assume that selection is operative. As yet, we do not know enough about the prevalence and integration of these two types of change, or how we can establish their relative importance in specific cases of regional and local economic evolution.

Complexity economics has undoubtedly proved a fruitful source of ideas, but to what extent does it also provide the basis of a coherent ontological conception of regional economic evolution that could be developed by economic geographers? Such a theory would have to be based on the premise that regional economic evolution is driven by advances in knowledge and that knowledge consists of rules or connections between ideas. The geography of knowledge is crucial to understanding the rate of macroeconomic growth as economic changes are the outcome of the relative balance between forces producing innovation, new knowledge and new variety, and forces leading to the disappearance of this variety through selection and the ageing of knowledge (Ramlogan and Metcalfe, 2006). This process is uneven across sectors and spaces:

> Growth does not occur without the continual emergence of innovation and the persistent changes in the relative importance of products, methods of production, firms, industries, regions and whole economies, that adaptation to innovation implies, and these changes in structure are a consequence and a cause of the growth of knowledge. (Ramlogan and Metcalfe, 2006, p. 133)

Dopfer and Potts (2004a) argue for a form of evolutionary realism that explains knowledge as evolutionary dynamics among systems of generic economic rules. The *micro* level of analysis refers to individuals' carrying of rules and actualisation of these rules, while the *macro* level is the population or deep structure of meso-rules that defines how rules coordinate with each other and fit together (Dopfer et al., 2004). At the *meso* level generic rules undergo phases of origination, diffusion and adaptation, and retention and replication.[21] The emergence of a new rule disrupts the coordinated structure and produces a period of de-coordination in actualizations. As the new rule moves through diffusion and retention phases, re-coordination occurs as a new division of labour, possibly involving regional and industrial organization, stabilises. Dopfer et al. (2004) argue that rules and their actualisations form 'meso units' that are the dynamical building blocks of an economic system:

> Work on industrial districts, regional knowledge clusters, learning regions, inter-firm organisation, national innovation systems, networks with weak or strong ties, or technical support communities all falls under the heading of meso economics from the evolutionary perspective. (Dopfer et al., p. 268)[22]

In this view an analysis of regional economic change depends on understanding how generic rules, which are composed of knowledge connections, emerge and are actualised and institutionalised in particular regions. While this view has considerable potential, the precise meaning and content of such rules seems to require much further clarification and illustration.

This emerging theoretical approach certainly has important implications for economic geography. First, it implies that there are strong reasons to re-examine and further develop the network perspectives deployed in parts of geography (for example, Castells, 2000; Taylor, 2003). It raises the question of whether the evolution of networks and connections can be analysed in terms of changes in underlying generative rules. Second, it suggests that the chreodic evolution of knowledge is the main reason why regional economies and local economies show forms of path dependence (see Loasby, 2001). Beinhocker (2006) proposes that human decision making is guided mainly by inductive reasoning, so that condition–action (if–then) rules used in the past will be applied in uncertain environments, and experienced mental models become resistant to change. This, he contends, explains the inertia and elimination of established businesses. In this view, the roots of path dependence in industrial clusters and regional economies lie in human decision-making. At the same time, inertia is never inevitable because of the diverse ways in which knowledge can change. As Ramlogan and Metcalfe (2006) argue, innovation arises in combinatorial knowledge and that this has no rest points but constantly undergoes chreodic change: 'It can undergo subtle changes as information percolates across networks of relationships or it can undergo sweeping changes that take understanding into entirely new dimensions. To this degree understanding is unstable' (2006, p. 130). The origins of regional path creation may well lie in this instability.

Third, this knowledge-based perspective also implies that the most effective complex systems are those that balance inertia and innovation. It is argued that if all the knowledge and beliefs in an organisation are changing rapidly and simultaneously then it proves impossible to exploit innovations. Complex organisations that have an intermediate mixture of inertia and innovativeness appear to have higher fitness (Hodgson and Knudsen, 2006b). In Ramlogan and Metcalfe's words: 'If beliefs are too fluid, order will descend into chaos; if beliefs are too rigid then order descends into lifeless equilibrium' (2006, p. 118). The degree to which this also applies to regional and local economies is an important research question. Are the most successful economies those that show the highest rates of innovation, and highest rates of global and local search, or are they those that are able to apply new innovations within relatively stable knowledge structures?

Fourth, the complexity paradigm also highlights some important challenges for contemporary economic geography relating to the importance of markets and personal knowledge. In similar fashion to many cultural economic geographers, complexity economists have recognised that knowledge exchange and connections fundamentally depend on institutions and patterns of institutionalization. Institutional rules store and communicate information and embody socially shared beliefs and understandings (Ramlogan and Metcalfe, 2006). However, in contrast to much recent economic geography, the complexity paradigm argues that markets play an essential and central role in coordinating knowledge and driving the evolution of knowledge. They are a central and distributed form of innovation system in modern capitalism. The growth of knowledge and markets are mutually reinforcing processes as markets do not simply match but continually reshape supply and demand. Selection processes are essentially market based as they distinguish reliable from less reliable knowledge connections and are a means of destroying some connections in favour of others. Markets also contain a great variability of types of linkage and many fluxes of connections, some of which are stabilised. In Potts' words: 'the market process is an experiment in knowledge: the creation of unforeseen compounds out of ephemeral elements that become obvious only after the event' (2001, p. 422). This view implies that economic geography's recent focus on 'untraded interdependencies', while certainly valuable, is missing or downplaying the most important cause and condition of the evolution of economic knowledge, namely markets. It surely implies that we should devote more attention to markets in regional and local economies both as experimental spaces and as mechanisms shaping production.

Finally, the paradigm implies that restless regional and local economies with dynamic rates of knowledge evolution will bring together two things. First they will have high knowledge heterogeneity as a result of having individuals with specialised and idiosyncratic knowledge that can be actualised as non-average behaviour. Second, such economies require high rates of information flow as these are necessary for the emergence and sustainability of social understanding. According to Metcalfe and Ramlogan (2005), shared social understanding in turn allows the growth of new idiosyncratic and specialised knowledge through the exchange and combination of information and also allows personal knowledge to be tested and put to use. Another implied challenge for economic geography then is not only to understand how institutional contexts shape collective understandings, but also to give much greater attention to the growth of personal knowledge and beliefs about economic action as these act as basic engines of economic change.

7. Some conclusions: fragments of a research agenda?

Our aim in this discursive chapter has been to begin to explore the scope and limits of 'complexity thinking' in evolutionary economic geography. It is clear that there are several key issues that require close attention in any research agenda aimed at constructing a complexity-based evolutionary approach to the subject. Essletzbichler and Rigby (2007) argue that economic geographers need to move beyond applying evolutionary ideas and concepts in an ad hoc manner, and develop a more general theory of economic evolution. In our view, complexity economics does not, as yet, provide the basis of such a theory, since there is no well defined or universally accepted complexity theory as such that economic geographers can simply turn to and apply to their own set of empirical issues and concerns. Rather, what exists is a series of generic notions about the characteristics and behaviour of complex systems, and most of these notions have their origin in the study of physical, chemical and biological systems. These do indeed provide some interesting potential concepts and ideas for thinking about socio-economic systems, and about the economic landscape and its evolution. Thus the notions of self-organisation, emergence and adaptation resonate closely with questions about how the spatial structure of an economy emerges and changes; about how regional and urban economies rise and fall in relative prosperity; about why some regional and urban economies appear more adaptable than others over time to shifts in technology, markets, policy regimes and the like; about why certain industries and technologies develop in particular geographical areas but not others; and about how the various spatial networks of economic relationships and flows form and evolve. In this sense we believe complexity thinking could make a valuable contribution to the construction of an evolutionary economic geography. But as in the case of the use of neo-Darwinian metaphors and ideas, the abduction of complexity ideas into economic geography is not at all straightforward.

One of the basic issues, we would argue, is ontological. The approach taken thus far in thinking about the economic landscape as a complex system has been what Perona would call 'theoretic', and what we would label 'scientific ontological' or even methodological; that is to say, it has been based on the use of formal models of complex systems drawn from the natural sciences to make claims about the nature and dynamics of (abstract) economic landscapes, as exemplified by the work of Arthur, Krugman and others. The problem with this approach is that the initiating premises, assumptions and processes are restricted to the content of the specific scientific models used, and these may not be adequate or realistic in relation to actually occurring, as opposed to hypothetical abstract and model-generated, economic landscapes. In such approaches, the characteristics of the economic landscape are thus those of the model, rather than those that actually exist.

Our argument here is that we need to look beyond the increasingly dominant modelling paradigms associated with complexity, including the functional development of appropriate computational architectures (such as multi-agent models and dynamical systems models), to a more philosophically inclined social-ontological approach. What precisely does it mean to talk of the economic landscape as a complex system? In what sense is the economic landscape a meaningful complex system to which the concepts of complexity thinking can be meaningfully applied? What does connectivity mean and how do we distinguish partial from strong connections? These are difficult questions. To be analytically useful, complexity is not something that just bolts on to or can be blended with an existing conceptual/theoretical framework to add a 'complexity perspective' or 'evolutionary perspective'. Nor is it sufficient to invoke the terminology and concepts of complexity science without thinking through what these concepts are being applied to, and what they mean in an economic-geographical context.

We take the view that if 'the economy' is indeed a complex system, its complexity arises in large part precisely because it is spatially distributed and spatially embedded. But this then behoves us to specify how this spatiality relates to complexity, both theoretically and empirically. We might adopt the terminology of Dopfer and Potts (2004b) and say that geography contributes to complexity because it increases the modularity, hierarchic depth and openness of the economy. The economy is modular because it is made up of a very large number of functional (sub)systems that are connected to (and interact with) one another (households, firms, institutions, states, and so on). The complexity of the economy is also a result of its hierarchic depth. That is, each (sub)system is itself complex in its own way, and simultaneously both made up of, and a component of, other systems: every part is a whole and every whole is a part. The economy is open in the sense that these systems interact via innumerable networks of connections and associations, and these connections can change. But exactly how does geographic space influence the modularity, hierarchic depth and openness of the economy?

From a complexity perspective, the spatial structure and organisation of the economy – the presence of features such as industrial districts, business clusters, cities, regional agglomerations, networks and the like – are to be understood as emergent properties of an economy, the unplanned meso-outcomes of the individual actions and behaviours of numerous individual economic agents (households, firms, institutions of various kinds, governments). Are such emergent spatial structures and features merely outcomes? Or are they themselves complex (sub)systems? Or are they just part of the 'environment' within which households, firms and other economic agents exist and function? Economic geographers have spent considerable effort demonstrating that spatial structures such as clusters, regional high-tech agglomerations, cities and the like are the source of a host of externalities and spillovers that influence, shape and regulate the behaviour of individual agents located within them (and indeed beyond). In this sense, such spatial-economic structures are not only examples of 'first-order' emergence (unintended spatial outcomes of myriads of micro-actions) but also of 'second-order' emergence (whereby these same meso-level spatial structures and arrangements – in conjunction with macro-level processes – feed back to influence micro-level behaviours and actions). Explicating these spatially emergent and spatially embedded systems of 'upward and downward causation' (Hodgson, 2004), and their multi-scalar operation and manifestation, would seem to us to be a critical task for a complexity-based evolutionary economic geography research agenda.

Likewise, how geographic-economic space both shapes and is shaped by the growth and transformation of knowledge are key ingredients of any complexity-based evolutionary economic geography. As we have seen, the most influential complexity economists argue that it is knowledge and its adaptation that makes the economy a complex system. But we know that the spatial localisation and agglomeration of economic agents is itself a major stimulus to the creation and circulation of knowledge: clusters, industrial districts and cities are quintessentially 'knowledge communities' (Loasby, 1998; Maskell, 2001; Pinch et al., 2003). In this sense, the creation of new knowledge – the engine of economic growth – is a spatially emergent effect, which then becomes part of the properties of economic agents (see Plummer and Sheppard, 2006). New knowledge (and innovation) typically emerges on a small scale in local contexts. But some of this new knowledge has the capacity to stimulate widespread adoption and large-scale transformations of the economic landscape. Conceptualising the role of geographic space in stimulating and conditioning the emergence, diffusion and adoption of economic novelty is to our mind a fundamental research task confronting evolutionary economic geographers. A complexity approach necessarily focuses attention on the co-evolution of knowledge and the economic landscape.

These are but some of the fragments of what is obviously a much larger research agenda. Whether and how far complexity thinking can help inform the construction of an evolutionary economic geography is as yet an open question. And as we have argued, complexity economics is itself underdeveloped. Yet, a complexity approach does seem to resonate with some of the central concerns of evolutionary economic geography. We know that the economic landscape is a highly complicated system, but it is also complex, in the specific sense of the term developed in this chapter: it is an open, highly interconnected, self-organising, emergent and adaptive system. The task is to construct an ontologically defensible framework based on this conception.

Notes

1. The original Nelson–Winter draws on Lamarckian rather than Darwinian evolutionary ideas. Although the processes of selection, mutation and inheritance are invoked not as biological metaphors but as real economic processes, they nevertheless explicitly acknowledge that there is an analogy between bio-genetic process and firm dynamics.
2. Note that we are not arguing against the use of analogies and metaphors per se. In fact, virtually all explanatory accounts make use of both. What matters of course is the relevance and appropriateness of the analogies or metaphors that are used. In this context, it is curious that Newtonian mechanistic analogies and metaphors should have dominated mainstream neoclassical economics for so long when they are clearly at odds with how actual socio-economic systems are constituted and develop.
3. Elsewhere in the *Principles of Economics*, Marshall acknowledges his debt to the writings of Herbert Spencer, and goes on to emphasise that 'biological conceptions are more complex than those of mechanics' (1930, p. xiv).

4. Sawyer (2005) argues that the interest in complex dynamical or adaptive systems represents the third wave of social systems theory. The first wave was structural functionalism, and general systems theory the second.
5. In their contribution, Harrison et al. (2006) explore the scope for using the ideas of complexity theory as a common basis for bridging human and physical geography.
6. It should be noted that Krugman's excursion into 'complexity theory' is limited in scope and makes no links with the ideas and concepts of evolutionary economics.
7. Indeed, 'complexity theory' is often portrayed as, and defined in terms of, the mathematics of non-linear dynamic stochastic systems, particularly physico-chemical systems.
8. Even in the physical and natural world, mathematical models typically fail to convey the precise nature of the specific *processes* at work.
9. A complex system can simply be chaotic, and is not necessarily self-organising. A complex adaptive system, however, is necessarily self-organising.
10. Indeed, there have also been attempts to trace embryonic 'complexity thinking' ideas in the history of economic thought, and to reinterpret the work of previous major economists – from Marx to Marshall – from a 'complexity perspective' (see Colander, 2000b).
11. Perona takes these terms directly from Lawson:

    If I can use the term *theoretic* to denote the quality of being a feature of a model and the term *ontic* to denote the quality of being features of the world the economist presumes to illuminate, a more succinct way of describing the problem that arises through the prioritisation of the modeling orientation is a conflation of the theoretic and ontic, with the latter reduced to the former. (Lawson, 2003, p. 4; emphasis in original)

12. This leads Potts to conclude that graph theory could be a useful methodology to deal with economic complexity at the representational level.
13. For example, Batten writes that 'The footprints of power laws can be found everywhere' (2001, p. 96).
14. In Foster's (2005) typology, 'first-order' complexity refers to the imposition of energy onto chemical elements such that patterns form in the dissipation of energy, while 'second-order' complexity refers to reception of information that is translated into a knowledge structure that permits control over the acquisition of energy. This is the level of complexity operating in the biological domain.
15. Schnabl et al. (1999) argue that input–output matrices reveal that the Queensland economy has become more interconnected and coordinated over time through market relations and has followed a coherent self-organising developmental path. Yet this surely reveals only one aspect of coordination.
16. This is indeed a general criticism that can be levelled at the idea of autopoiesis when applied to social systems (see Mingers, 2002).
17. The formation of institutions can also be the result of co-evolution or self-organisation, and is thus not exogenous to the system under study. Institutions might be assumed as exogenous in the short run, but themselves become endogenous in the longer term, as they change and transform as the economy itself evolves (see Martin and Sunley, 2006).
18. Potts expressly links 'lock-in' to low connectivity (ordered) systems:

    In the ordered regime, the particular configuration of connections locked into (descriptive of preferences, technology, institutions) may or may not be optimal in the sense that other combinations may be better, but the system has no internal mechanism to change to these states. (2000, p. 90)

19. Later and perhaps inconsistently, however, Beinhocker swallows a reductionist version of Putnam and argues that economies with more connections in the form of social capital have more trust and grow more quickly.
20. Foster (1997) also appears to argue that a complexity view focuses on selection via changes in system boundary conditions rather than by microscopic dynamics.
21. Beinhocker (2006) also argues that there are co-evolutionary dynamics that simultaneously shape social technologies, physical technologies and business plans. Such plans are carried by businesses that act as replicators. These plans are built from amalgamations of modules that themselves are combinations of atomistic physical and social technologies. He defines modules as 'a component of a business plan that has provided in the past, or could provide in the future, a basis for differential selection between businesses in a competitive environment' (p. 283). Is this a rehearsal of, or an advance on, Nelson and Winter's neo-Darwinian evolutionism?
22. In an interesting paper, Mehier and Brette (2005) seek to use Dopfer's *micro–meso–macro* ontology to formulate some hypotheses concerning the life-cycle of clusters. They argue that Dopfer's ontology equates to Veblen's institutional framework, whereby micro relates to individual habits of action and

thought, meso to institutions and 'institutional logics', and macro to the 'cultural complex' of society. They go on to suggest that a cluster can be considered to be a spatial institutionalisation of an agglomeration rule, and hence amenable to analysis in terms of Dopfer's idea of meso trajectories (origination, diffusion and retention of a novel rule in an economic system).

References

Adami, C. (2002), 'What is complexity?', *BioEssays*, **24**, 1085–94.
Allen, P.M. (1982), 'Self-organisation in the urban system', in W. Shieve and P.M. Allen (eds), *Self Organisation and Dissipative Structures: Applications in the Physical and Social Sciences*, Austin: University of Texas Press, pp. 132–58.
Allen, P.M. (1985), 'Towards a science of complex systems', in S. Aida et al. (eds), *The Science and Praxis of Complexity*, Tokyo: United Nations University, pp. 268–97.
Allen, P.M. (1992), 'Modelling evolutionary and complex systems', *World Futures*, **34**, 105–23.
Anderson, P.W., K.J. Arrow and D. Pines (eds) (1988), *The Economy as an Evolving, Complex System*, Reading, MA: Addison-Wesley.
Arthur, W.B. (1988), 'Self-reinforcing mechanisms in economics', in P. Anderson, K. Arrow and D. Pines (eds), *The Economy as an Evolving, Complex System*, Reading, MA: Addison-Wesley, pp. 9–31.
Arthur, W.B. (1989), 'Competing technologies, increasing returns, and "lock-in" by historical events', *Economic Journal*, **99**, 116–31.
Arthur, W.B. (1994), *Increasing Returns and Path Dependence in the Economy*, Michigan: Michigan University Press.
Arthur, W.B., S. Durlauf and D. Lane (eds) (1997), *The Economy as a Complex Evolving System, II*, Reading, MA: Perseus Books.
Auyang, S. (1998), *Foundations of Complex-System Theories in Economics, Evolutionary Biology and Statistical Physics*, Cambridge: Cambridge University Press.
Bak, P. (1996), *How Nature Works: The Science of Self-Organised Criticality*, New York: Springer Verlag.
Bathelt, H. and J.S. Boggs (2003), 'Towards a reconceptualization of regional development paths: is Leipzig's media cluster a continuation of or a rupture with the past?', *Economic Geography*, **79** (2), 265–93.
Batten, D.F. (2001), 'Complex landscapes of spatial interaction', *Annals of Regional Science*, **35**, 81–111.
Batty, M., J. Barros and S.A. Junior (2004), 'Cities: continuity, transformation and emergence', Working Paper 72, CASA UCL, London.
Beinhocker, E.D. (2006), *The Origin of Wealth: Evolution, Complexity and the Radical Remaking of Economics*, London: Random House.
Bentley, R.A. and H.D.G. Maschner (2001), 'Stylistic change as a self-organised critical phenomenon: an archaeological study in complexity', *Journal of Archaeological Method and Research*, **8**, 35–66.
Bentley, R.A. and H.D.G. Maschner (2003), *Complex Systems and Archaeology: Empirical and Theoretical Applications*, Salt Lake City: University of Utah Press.
Boschma, R.A. and K. Frenken (2003), 'Evolutionary economics and industrial location', mimeo, Department of Economic Geography, Utrecht University.
Boschma, R.A. and K. Frenken (2006), 'Why is economic geography not an evolutionary science?', *Journal of Economic Geography*, **6**, 273–302.
Boschma, R.A. and J.G. Lambooy (1999), 'Evolutionary economics and economic geography', *Journal of Evolutionary Economics*, **9** (4), 411–29.
Buchanan, M. (2004), 'It's the economy, stupid', *New Scientist*, **82**, 34–7.
Bullock, S. and D. Cliff (2004), 'Complexity and emergent behaviour in ICT', report commissioned by Foresight Programme, OST, Hewlett Packard.
Byrne, D.S. (1998), *Complexity Theory and the Social Sciences*, London: Routledge.
Castells, M. (2000), *The Rise of the Network Society: The Information Age*, Volume 1, Oxford: Blackwell.
Chattoe, E. (2006), 'Using evolutionary analogies in social science: two case studies', in A. Wimmer and R. Kössler (eds), *Understanding Change: Models, Methodologies and Metaphors*, Basingstoke: Palgrave Macmillan, pp. 89–95.
Cilliers P. (2001), 'Boundaries, hierarchies and networks in complex systems', *International Journal of Innovation Management*, **5** (2), 135–47.
Cilliers, P. (2005a), 'Complexity, deconstruction and relativism', *Theory, Culture and Society*, **22** (5), 255–67.
Cilliers, P. (2005b), 'Knowledge, limits and boundaries', *Futures*, **37**, 605–13.
Colander, D.A. (ed) (2000a), *The Complexity Vision and the Teaching of Economics*, Cheltenham, UK and Northampton, MA, USA: Edward Elgar.
Colander, D.A. (ed.) (2000b), *Complexity and the History of Economic Thought*, London: Routledge.
Dopfer, K., J. Foster and J. Potts (2004), 'Micro–meso–macro', *Journal of Evolutionary Economics*, **14**, 263–79.

Dopfer, K. and J. Potts (2004a), 'Evolutionary realism: a new ontology for economics', *Journal of Economic Methodology*, **11** (2), 195–212.

Dopfer, K. and J. Potts (2004b), 'Evolutionary foundations of economics', in J.S. Metcalfe and J. Foster (eds), *Evolution and Economic Complexity*, Cheltenham, UK and Northampton, MA, USA: Edward Elgar, pp. 3–23.

Essletzbichler, J. and D. Rigby (2004), 'Competition, variety and the geography of technology evolution', *Tijdschrift voor Economische en Sociale Geografie*, **96** (1), 48–62.

Essletzbichler, J. and D. Rigby (2007), 'Exploring evolutionary economic geography, *Journal of Economic Geography*, **7** (5), 537–48.

Frenken, K. and R. Boschma (2007), 'A theoretical framework for evolutionary economic geography: industrial dynamics and urban growth as a branching process', *Journal of Economic Geography*, **7** (5), 635–49.

Foster, J. (1993), 'Economics and the self organization approach: Alfred Marshall revisited', *Economic Journal*, **103**, 975–91.

Foster, J. (1994), 'The self organisation approach in economics', in S. Burley and J. Foster (eds), *Economics and Thermodynamics: New Perspectives on Economic Analysis*, Boston: Kluwer, pp. 183–202.

Foster, J. (1997), 'The analytical foundations of evolutionary economics: from biological analogy to economic self organization', *Structural Change and Economic Dynamics*, **8**, 427–51.

Foster, J. (2000), 'Competitive selection, self-organisation and Joseph A. Schumpeter', *Journal of Evolutionary Economics*, **10**, 311–28.

Foster, J. (2004), 'Why is economics not a complex systems science?', Discussion Paper 336, School of Economics, University of Queensland.

Foster, J. (2005), 'From simplistic to complex systems in economics', *Cambridge Journal of Economics*, **29**, 873–92.

Frenken, K. and R. Boschma (2007), 'A theoretical framework for evolutionary economic geography: industrial dynamics and urban growth as a branching process', *Journal of Economic Geography*, **7** (5), 635–49.

Frenken, K. and A. Nuvolari (2004), 'The early development of the steam engine: an evolutionary interpretation using complexity theory', *Industrial and Corporate Change*, **13** (2), 419–50.

Garnsey, E. (1998), 'The genesis of the high technology milieu: a study in complexity', *International Journal of Urban and Regional Research*, **22** (3), 361–77.

Gattrell, A.C. (2005), 'Complexity theories and geographies of health: a critical assessment', *Social Science and Medicine*, **60**, 2661–71.

Gilbert, N. (2002), 'Varieties of emergence', mimeo, Centre for Research on Simulation in the Social Sciences, University of Surrey.

Harrison, S. D. Massey and K. Richards (2006), 'Complexity and emergence (another conversation)', *Area*, **38**, 465–71.

Hassink, R. (2005), 'How to unlock regional economies from path dependency? From learning region to learning cluster', *European Planning Studies*, **13** (4), 521–35.

Heylighen, F. (1999), 'The science of self-organization and adaptivity', in L.D. Kiel (ed.), 'Knowledge management, organizational intelligence and learning, and complexity', in *The Encyclopaedia of Life Support Systems*, Oxford: EOLSS Publishers.

Hodgson, G. (1993), *Economics and Evolution: Bringing Life Back into Economics*, Cambridge: Polity Press.

Hodgson, G. (2004), *The Evolution of Institutional Economics: Agency*. *Structure and Darwinism in American Institutionalism*, New York: Routledge.

Hodgson, G. and T. Knudsen (2006a), 'Why we need a generalized Darwinism, and why generalized Darwinism is not enough', *Journal of Economic Behaviour and Organization*, **61**, 1–19.

Hodgson, G. and T. Knudsen (2006b), 'Balancing inertia, innovation, and imitation in complex environments', *Journal of Economic Issues*, **XL** (2), 287–95.

Holland, J. (1992), *Adaptation in Natural and Artificial Systems: An Introductory Analysis with Applications to Biology, Control and Artificial Intelligence*, Cambridge, MA: MIT Press.

Holland, J. (1995), *Hidden Order: How Adaptation Builds Complexity*, New York: Helix Books.

Horgan, J. (1995), 'From complexity to perplexity', *Scientific American*, **272**, 104–9.

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

Kaufman, S. (1995), *At Home in the Universe: The Search for the Laws of Self Organisation and Complexity*, New York: Oxford University Press.

Krugman, P. (1994), 'Complex landscapes in economic geography', *American Economic Review*, **84**, 412–16.

Krugman, P. (1996), *The Self-Organising Economy*, Cambridge, MA: MIT Press.

Krugman, P. (1997), 'How the economy organises itself in space: a survey of the new economic geography', in W. Arthur, S. Durlauf and D. Lane (eds), *The Economy as a Complex Evolving System, II*, pp. 239–62.

Lambooy, J.G. and R.A. Boschma (2001), 'Evolutionary economics and regional policy', *Annals of Regional Science*, **35**, 113–31.
Lawson, T. (2003), 'The (confused) state of equilibrium analysis in modern economics: an (ontological) explanation', Paper presented at the London History and Philosophy of Science Meetings, LSE-CPNSS, London, UK.
Lindsay, V.J. (2005), 'The development of international industry clusters: a complexity theory approach', *Journal of International Entrepreneurship*, **3**, 71–97.
Lissack, M.F. (1999), 'Complexity: the science, its vocabulary, and its relation to organizations', *Emergence*, **1**, 110–26.
Loasby, B. (1998), 'Industrial districts as knowledge communities', in M. Bellett and C. L'Harmet (eds), *Industry, Space and Competition*, Cheltenham, UK and Northampton, MA, USA: Edward Elgar, pp. 70–85.
Loasby, B. (2001), 'Industrial dynamics: why connections matter', DRUID Working Paper No. 01–09.
McGlade, J. (2006), 'Ecohistorical regimes and *la Longue Durée*: an approach to mapping long-term societal change', in E. Garnsey and J. McGlade (eds), *Complexity and Co-Evolution: Continuity and Change in Socio-Economic Systems*, Cheltenham, UK and Northampton, MA, USA: Edward Elgar, pp. 77–114.
McGlade, J. and E. Garnsey (2006), 'Introduction: the nature of complexity', in E. Garnsey and J. McGlade (eds), *Complexity and Co-Evolution: Continuity and Change in Socio-Economic Systems*, Cheltenham, UK and Northampton, MA, USA: Edward Elgar, pp. 1–21.
McKelvey, B. (1999), 'Avoiding complexity catastrophe in coevolutionary pockets: strategies for rugged landscapes', *Organization Science*, **10** (3), 294–321.
Manson, S. and D. O'Sullivan (2006), 'Complexity theory in the study of space and place', *Environment and Planning A*, **38**, 677–92.
Markose, S.M. (2005), 'Computability and evolutionary complexity: markets as complex adaptive systems', *The Economic Journal*, **115**, 159–92.
Marshall, A. (1930), *Principles of Economics*, London: Macmillan.
Martin, R.L. and P.J. Sunley (2006), 'Path dependence and regional economic evolution', *Journal of Economic Geography*, **6**, 395–435.
Maskell, P. (2001), 'Towards a knowledge-based theory of the geographical cluster', *Industrial and Corporate Change*, **10**, 919–41.
Mehier, M. and O. Brette (2005), 'Veblen's evolutionary economics revisited through the micro–meso–macro framework: the stakes for the analysis of clusters of innovation', STOICA Working Paper 2005–09–05, University of Lyon.
Metcalfe, J.S. (1997), 'Evolutionary concepts in relation to evolutionary economics', Queensland University Discussion Paper, 226.
Metcalfe, J.S. (1998), *Evolutionary Economics and Creative Destruction*, London: Routledge.
Metcalfe, J.S. and J. Foster (eds) (2004), *Evolution and Economic Complexity*, Cheltenham, UK and Northampton, MA, USA: Edward Elgar.
Metcalfe, J.S and R. Ramlogan (2005), 'Limits to the economy of knowledge and knowledge of the economy', *Futures*, **37**, 655–74.
Metcalfe, J.S., J. Foster and R. Ramlogan (2006), 'Adaptive economic growth', *Cambridge Journal of Economics*, **30**, 7–32.
Mingers, J. (2002), 'Can social systems be autopoietic? Assessing Luhmann's theory', *The Sociological Review 50*, **2**, 278–99.
Nelson, R.R. and S. Winter (1984), *An Evolutionary Theory of Economic Change*, Cambridge, MA: Belknap Harvard University Press.
Nicolis, G. and I. Prigogine (1977), *Self Organisation in Non-Equilibrium Systems*, New York: John Wiley.
Nicolis, G. and I. Prigogine (1989), *Exploring Complexity*, New York: W.H. Freeman and Co.
O'Sullivan, D., P. Manson, J.P. Messina and T.W. Crawford (2006), 'Space, place, and complexity science', *Environment and Planning A*, **38**, 611–17.
Pavard, B. and J. Dugdale (2000), *An Introduction to Complexity in Social Science*, GRIC-IRIT, Toulouse, available at: http://www.irit.fr/COSI/training/complexity-tutorial/complexity-tutorial.htm
Penrose, E. (1952), 'Biological analogies in the theory of the firm', *American Economic Review*, **XLII**, 804–19.
Perona, E. (2004), 'The confused state of complexity economics: an ontological explanation', Paper presented at the 'Complexity and Economic Policy Conference', University of Salerno.
Pinch, S., N. Henry, M. Jenkins and S. Tallman (2003), 'From "Industrial districts" to "knowledge clusters": a model of knowledge dissemination and competitive advantage industrial agglomerations', *Journal of Economic Geography*, **3**, 373–88.
Plummer, P. and E. Sheppard (2006), 'Geography matters: agency, structures and dynamics at the intersection of economics and geography', *Journal of Economic Geography*, **6**, 619–38.
Porter, M. (1998), *On Competition*, Cambridge, MA: Harvard Business School Press.

Potts, J. (2000), *The New Evolutionary Micro-economics: Complexity, Competence and Adaptive Behaviour*, Cheltenham, UK and Northampton, MA, USA: Edward Elgar.
Potts, J. (2001), 'Knowledge and markets', *Journal of Evolutionary Economics*, **11**, 413–31.
Ramlogan, R. and S. Metcalfe (2006), 'Restless capitalism: a complexity perspective on modern capitalism', in E. Garnsey and J. McGlade (eds), *Complexity and Co-Evolution*, Cheltenham, UK and Northampton, MA, USA: Edward Elgar, pp. 115–46.
Rigby, D.L. and J. Essletzbichler (1997), 'Evolution, process variety, and regional trajectories of technological change in U.S. manufacturing', *Economic Geography*, **73**, 269–83.
Robson, A.J. (2005), 'Complex evolutionary systems and the Red Queen', *The Economic Journal*, **115** (June), F211–F224.
Rosenau, J.N. (1995), 'Governance in the 21st century', *Global Governance*, **1**, 13–44.
Rosser, J.B. (2004), *Complexity in Economics*, Cheltenham, UK and Northampton, MA, USA: Edward Elgar.
Sawyer, R.K. (2005), *Social Emergence: Societies as Complex Systems*, Cambridge: Cambridge University Press.
Schenck, K.E. (2003), *Economic Institutions and Complexity: Structures, Interactions and Emergent Properties*, Cheltenham, UK and Northampton, MA, USA: Edward Elgar.
Schnabl, H., G. West, J. Foster and A. Cook (1999), 'A new approach to identifying structural development in economic systems: the case of the Queensland economy', *Australian Economic Papers* (March), 64–78.
Schweitzer, F. (ed.) (1997), *Self Organisation of Complex Structures: From Individual to Collective Dynamics*, Amsterdam: Gordon and Breach.
Stacey, R.D., D. Griffin and P. Shaw (2000), *Complexity and Management: Fad or Radical Challenge to Systems Thinking?*, London: Routledge.
Taylor, P.J. (2003), *World City Network: A Global Urban Analysis*, London: Routledge.
Thrift, N. (1999), 'The place of complexity', *Theory, Culture and Society*, **16** (3), 31–69.
Viskovatoff, A. (2000), 'Will complexity turn economics into sociology?', in D.A. Colander (ed.), *Complexity and the History of Economic Thought*, London: Routledge, pp. 129–54.
Wakeley, T. (2002), 'Review of *The New Evolutionary Microeconomics: Complexity, Competence and Adaptive Behaviour* by Jason Potts, Cheltenham, Edward Elgar', *Journal of Economic Psychology*, **23**, 279–86.
Wimmer, A. (2006), 'Models, methodologies and metaphors on the move', in A. Wimmer and R. Kössler (eds), *Understanding Change: Models, Methodologies and Metaphors*, Basingstoke: Palgrave Macmillan, pp. 1–36.
Witt, U. (1999), 'Bioeconomics as economics from a Darwinian Perspective', *Journal of Bioeconomics*, **1**, 19–34.
Witt, U. (2003), *The Evolving Economy: Essays on the Evolutionary Approach to Economics*, Cheltenham, UK and Northampton, MA, USA: Edward Elgar.
Wolfram, S. (2002), *A New Kind of Science*, Champaign, IL: Wolfram Media.

5 The spatial evolution of innovation networks: a proximity perspective

Ron Boschma and Koen Frenken

1. Introduction

The role of networks in innovation processes has become a key research area in the field of innovation studies over the last decade and a half (Freeman, 1991; Hagedoorn, 2002; Powell et al., 1996). Not surprisingly, the rapid increase in the number of studies on innovation networks in an inter-disciplinary field as innovation studies has led to a great variety of theories, concepts and methodologies (Ozman, 2009). Only recently, geographers have jumped on the empirical study of the spatial dimensions of networks in innovation processes, following the vast literature on national and regional innovation systems developed in the 1990s (Ter Wal and Boschma, 2009a). Despite this attention, network analysis is still underdeveloped in the geography of innovation. This is also true for an evolutionary approach to this topic, although attempts have been undertaken (see e.g. Giuliani, 2007; Glückler, 2007; and Chapters 16, 17, 12 and 14 in this volume by Breschi et al., Cantner and Graf, Giuliani, and Glückler).

Our aim is to propose an evolutionary perspective on the geography of network formation that is firmly grounded in a dynamic proximity framework. Doing so, we link the emerging literatures on network and proximity dynamics. The study on network evolution is still in a premature phase (Powell et al., 2005), though considered crucial for the development of an evolutionary perspective on the geography of innovation networks. Following Boschma (2005), we present various forms of proximity as alternative driving forces behind network formation. In doing so, we root the proximity concept in an evolutionary approach to the geography of innovation networks. In this chapter, we discuss three topics.

The first topic focuses on explaining the structure of networks. For instance, why are some individuals or networks better connected than others? Do individuals and firms that are geographically proximate show a higher degree of connectivity? We use ideas obtained from the French school on proximity dynamics (e.g. Rallet, 1993; Rallet and Torre, 1999) to explain the formation of innovation networks. They state that other forms of proximity besides geographical proximity may facilitate interactive learning and innovation. In that context, Boschma (2005) has claimed that geographical proximity is neither a necessary nor a sufficient condition for learning and innovation. In this chapter, we will present the different forms of proximity as alternative forces driving network formation.

The second topic concentrates on explaining the effects of networks. For instance, does connectivity increase the innovativeness of organizations? And do local or non-local networks affect the performance of organizations (Bathelt et al., 2004)? While a high degree of proximity might be considered a prerequisite to make agents connected, when assessing the effects of network linkages, we argue that proximity between agents does not necessarily increase their innovative performance, and may possibly even harm it (Boschma, 2005; Broekel and Meder, 2008). We refer to this as the proximity paradox, and claim that it depends on the (optimal) level of proximity between agents whether their connection will lead to a higher level of innovative performance or not.

The third topic deals with the long-term dynamics of networks and the changing role of proximity dimensions in the formation and performance of innovation networks. For instance, do networks of innovation become less geographically proximate over time during the course of an industry lifecycle (e.g. Ter Wal, 2009)? And when does the evolution of a network structure show tendencies of path dependence, and why? We argue that the different proximities may induce path dependence in network evolution, and may cause retention in the local network (Glückler, 2007). We explain how local network retention might lead to regional lock-in. Last but not least, we argue that a dynamic network approach should account for that fact that the evolution of network structures may, in turn, affect the degree of the different forms of proximity (Menzel, 2008; Ter Wal and Boschma, 2009b).

2. Network structure and the proximity concept

A key question in (innovation) network research is to explain the presence or absence of network relations between organizations, or, more generally, the number or strength of relationships between actors in a network. The dependent variable is thus the bilateral relation. The main strategy to explain network structure, then, is to compare the similarity between actors that are linked with the similarity between actors that are not linked. For example, social networks are generally structured along the lines of gender, ethnicity, age and education as people have a bias to make friends of the same sex, ethnicity, age group and education level. Sociologists call similarity in attributes of nodes *homophily*, but we follow the terminology of *proximity*, which is more common in innovation studies and related areas with a focus on inter-organizational networks (Boschma, 2005; Knoben, 2007; Lagendijk and Oinas, 2005; Rallet, 1993; Rallet and Torre, 1999).

Though scholars differ in the definition of proximity and the number of proximity dimensions, we follow Boschma (2005) in his definition of five forms of proximity: cognitive, organizational, social, institutional and geographical proximity. In short, cognitive proximity indicates the extent to which two organizations share the same knowledge base; organizational proximity, the extent to which two organizations are under common hierarchical control; social proximity, the extent to which members of two organizations have friendly relationships; institutional proximity, the extent to which two organizations operate under the same institutions; and geographical proximity, the physical distance or travel time separating two organizations. These proximity dimensions are discussed below in more detail.

We believe the proximity concept is part and parcel of an evolutionary approach. In an evolutionary approach, firms innovate in areas close to their current cognitive capabilities along well-defined technological trajectories (Nelson and Winter, 1982). Their distinctive capabilities also constitute the primary determinants on partner selection in innovation networks. To exchange knowledge and develop innovations, networking firms tend to be close, yet complementary in cognitive/technological space. The concept of proximity in five dimensions as defined by Boschma (2005) can thus be regarded as an extension of the evolutionary approach that focuses on cognitive proximity primarily:

proximity is required in some (but not necessarily all) dimensions to get firms connected and to enable interactive learning and innovation among them.

Cognitive proximity

The effective transfer of knowledge and collaboration requires absorptive capacity to identify, interpret and exploit the new knowledge (Cohen and Levinthal, 1990; Nooteboom, 1999, 2000). For this reason, the capacity of actors or firms to absorb new knowledge requires cognitive proximity. That is, their own cognitive base should be close enough to the new knowledge in order to communicate, understand and process it successfully. With the notion of cognitive proximity, it is meant that people or firms sharing the same knowledge base and expertise are expected to learn more from each other than if cognitive distance is large. Nooteboom et al. (2007), among others, have demonstrated that cognitive proximity is indeed an important determinant in R&D alliances. It is also visible in patent citations, which have been considered as proxies for knowledge spillovers. For instance, Breschi and Lissoni (2006) found that most patent citations occur within the same 12 digit patent class. Studies focusing on a specific cluster observe that cluster firms perform different roles in knowledge networks because they differ in cognitive terms. Some firms act as hubs, while other cluster firms are poorly connected because they lack the capabilities to understand and exploit external knowledge (Boschma and Ter Wal, 2007; Giuliani and Bell, 2005; Morrison, 2008).

Organizational proximity

Boschma (2005) defined organizational proximity as the extent to which relations are shared in an organizational arrangement, either within an organization, or between organizations. It involves the rate of autonomy and control that can be exerted in organizational arrangements. A continuum is assumed, ranging from one extreme of ‘on the spot’ market, to informal relations between firms (e.g. interlocking corporate boards), to more formal organizational networks (e.g. a joint-venture, franchise), to the other extreme of a hierarchically organized firm (Williamson, 1985). As for cognitive proximity, organizational proximity is believed to be beneficial for establishing innovation networks, because they reduce uncertainty and opportunism. Strong control mechanisms are required to ensure ownership rights and sufficient rewards for own investments in new technology. Markets are poorly equipped to fulfil these tasks, because they tend to generate excessive transaction costs. In addition, formal contracting is almost impossible when it concerns complex and long-term research collaborations in which it is hard to determine and codify what activities will be undertaken, and what kinds of return will be generated (Nooteboom, 1999).

Social proximity

The notion of social proximity has its roots in the embeddedness literature (Granovetter, 1985; Uzzi, 1996). This literature indicates that economic relations are always embedded in a social context and that, in turn, social relations affect economic outcomes. Boschma (2005) defined social proximity in terms of socially embedded relations between agents at the micro-level. Relations between actors are socially embedded when they involve trust that is based on friendship, kinship and experience through repeated interaction. Such relationships carry information about potential partners and thereby increase the probability of organizations to engage in innovation networks. What is more, the perceived risk of conflict is also lower as social proximity adds to trust among organizations. Social proximity also plays a role in informal knowledge exchange between employees affiliated to different organizations. Breschi and Lissoni (2003, 2006) found that social connectedness between inventors played a significant role in knowledge spillovers. That is, social networks based on personal acquaintances as a result of common working experiences are important carriers of knowledge exchange based on reciprocity. Agrawal et al. (2006) point out that firms often connect because their employees used to work for the same organization in the past. These findings support the concepts of epistemic communities, invisible colleges and communities of practice. A particular mechanism in which social proximity plays a key role in the formation of new network relations is known as 'closure', which refers to ties that are created when two nodes are introduced to one another by a common third with whom both already have a network relation.

Institutional proximity

Whereas social proximity is defined in terms of socially embedded relations between agents at the micro-level, institutional proximity is associated with institutions at the macro-level. Both formal institutions (as laws) and informal institutions (like cultural norms and values) influence the extent and the way organizations coordinate their actions (Edquist and Johnson, 1997; Hall and Soskice, 2001; Hofstede, 1991). As such, institutions are enabling mechanisms that provide stable conditions for interactive learning. A classic study on institutional proximity has been an empirical study on the adoption of German machinery in Canadian firms (Gertler, 1995). The problems in using and maintaining the machinery could be related to different macro-institutions. In Germany with long-life employment and on-the-job training, employees had little difficulty in operating complex machinery, while in Canadian firms, with high turnover of personnel and little intra-firm training, employees had difficulty operating and maintaining the complex machinery. This example shows that inter-firm relationships are often hampered by a lack of institutional proximity between countries (Hall and Soskice, 2001). Another example of a lack of institutional proximity is in university–industry–government or 'triple helix' relationships (Etzkowitz and Leydesdorff, 2000), where different key actors operate in different institutional regimes.

Geographical proximity

The final dimension to be distinguished is geographical proximity. There is a strong claim that geographical proximity is a prime mover of network formation despite globalization, implying that a great deal of interactions still take place between agents that are geographically proximate (see e.g. Hoekman et al., 2009; Weterings, 2005). Once having defined the four other forms of proximity, geographical proximity can be defined in a restricted manner as the physical distance between actors in absolute (e.g. miles) or relative terms (e.g. travel time) (Boschma, 2005). Geographical proximity is beneficial for innovation as effective learning requires face-to-face interaction. Such interaction is easier (and cheaper) to organize when agents are co-located. The relationship between geographical proximity and co-location is not that straightforward though, because they do not necessarily mean the same thing. The need for geographical proximity (or better, face-to-face interactions) may be realized by temporary co-location (bringing agents together by means of fairs, conferences, business meetings, et cetera), instead of permanent co-location (Torre, 2008; Torre and Rallet, 2005). In sum, for analytical purposes, it is essential to define geographical proximity in such a restricted manner, and to isolate it from the other dimensions of proximity.

As proximity is an analytical concept, it offers many advantages in theoretical and empirical work explaining the (spatial) structure of networks.

First, one can extend the list of relevant proximity dimensions with any other dimension without changing the meaning of each dimension. For example, linguistic or ethnic proximity can be introduced. Thus, the proximity dimensions are analytically orthogonal even though many dimensions of proximity may empirically turn out to be correlated. Just to give one example, social proximity between two organizations is generally higher for geographically proximate organizations, because friendships are more easily established and maintained over short distances.

Second, by incorporating multiple proximity dimensions in an explanatory framework, one can test what forms of proximity are best able to explain the formation of networks. For example, many networks are geographically localized. One is then tempted to argue that this is the case because of transportation costs. However, if one explains the presence or absence of links using indicators of both geographical proximity and social proximity, one may find that networks are actually based on social proximity and not on geographical proximity. Yet, if firms with a high social proximity are often co-located, it can seem that geographical proximity is underlying the formation of networks. Thus, ideally, one takes into account as many proximity dimensions as possible as to control for all possible reasons that may underlie network formation between organizations.

Third, the analytical nature of proximity concept allows one to understand the interplay between different dimensions. In particular, one can expect proximity dimensions in innovation networks to be substitutes rather than complements (Boschma, 2005). That is to say, to establish a (successful) relation, one is in need of proximity in at least one dimension to manage the uncertainty involved. Being proximate in a second dimension, then, adds relatively little to the probability a link is formed, or the probability that the relation is successful. Making use of patent data, Singh (2005) found that geographical proximity is especially important in the establishment of interdisciplinary research collaboration (when cognitive proximity is low), while inventors working in the same field (i.e. cognitive proximity is high) collaborate on average over longer geographical distances. Making use of publication data, Ponds et al. (2007) found that geographical proximity is especially important in the establishment of university–industry–government relationships (i.e. institutional proximity is low) and less important in university–university collaboration where actors operate under the same institutions (i.e. institutional proximity is high). Agrawal et al. (2006) found that knowledge is transferred between firms in different locations (so geographical proximity is low) by employees that are socially linked because of a shared past. Breschi et al. (Chapter 16 in this volume) found similar results when analysing the social networks of US inventors who are mobile in space. Although inter-regional mobility of inventors is very low, the few inventors who did move between regions often maintained their ties with former co-inventors, providing a channel of knowledge diffusion to their prior location.

3. Network effects on organizational performance

The effect of networks on the performance of organizations is a second key question in (innovation) network research. Generally, the effects of having networks relations are positive. Most studies find a positive relationship between the number of network relations of a firm and its performance (see Ozman, 2009 for an overview). The same holds for informal social networks as, for example, evidenced by the finding that social networks between two people significantly increase the probability of knowledge spillovers (Breschi and Lissoni 2003, 2006). At the more aggregate level of regions in the European Union, the impact of collaboration networks on regional innovative performance has been analysed by explaining the number of patents by knowledge inputs weighted for the number of collaborations existing between the regions. The results show that the collaboration networks between regions indeed provide access for a region to the scientific knowledge in other regions (Hoekman et al., 2008; Maggioni et al., 2007).

The effect of networks on performance has been further elaborated by distinguishing between different types of knowledge. Sorenson et al. (2006) analysed US patent data and citation rates across proximate and distant actors on three dimensions of proximity: (1) social proximity (concerning distance between inventors in a network of patent collaborators); (2) geographical proximity (spatial distance between inventors); (3) organizational proximity (firm membership). They came to the conclusion that the advantages of being geographically proximate to some knowledge source depend crucially on the nature of the knowledge at hand. With respect to simple knowledge and very complex knowledge, the results show that more close actors are not in a more advantageous position, as compared to more distant actors. Simple knowledge flows equally to actors near and far, while complex knowledge is unlikely to diffuse, no matter how proximate actors are. With knowledge of moderate complexity, however, the outcomes show that more close actors are in a better position to benefit from knowledge diffusion, in contrast to more distant recipients.

However, these results should not be taken to mean that any network relation will have a positive effect. Each network relation comes at a cost, both in its establishment (search, negotiation) and its maintenance (conflict, monitoring). In the context of innovation networks, a particular risk in networking is the risk of involuntary knowledge spillovers through which valuable knowledge leaks to other organizations. Conflicts may arise as well. The main rationale of agents to share information and knowledge is that they expect such favours to be reciprocal. Once an agent persistently fails to reciprocate, the network linkages will become unstable, and will not deliver any positive effects. In addition to that, too much proximity between agents in networks may lead to lock-in situations (Boschma, 2005). Excess cognitive proximity reduces the scope for learning (Nooteboom, 2000). Two people or organizations with the same knowledge have little to exchange. Knowledge creation often requires dissimilar, complementary bodies of knowledge, especially in the context of radical innovations. Cognitive proximity also increases the risk of involuntary knowledge spillovers, especially when the new knowledge cannot be fully appropriated. With respect to social proximity, socially embedded relationships may lead to excess loyalty such that an agent puts their friends' interests before their own (Uzzi, 1996). Moreover, long-term relationships, or too much commitment may lock members of social networks into established ways of doing things, which may be harmful for learning.

Consequently, one ends up in a paradoxical situation. In section 2, we have gone at length to explain that a high degree of proximity is considered a prerequisite to make agents connected. However, when assessing the economic effects of networks, we argue that proximity between agents in networks does not necessarily increase their innovative performance, and may even harm it (Boschma, 2005; Broekel and Meder, 2008). We refer to this as the proximity paradox. When incorporating a proximity framework in network analysis, one should therefore make a distinction between the drivers of network formation on the one hand (in which the forms of proximity positively affect the establishment of networks), and the effects of network on innovative performance on the other hand (in which it is uncertain what the effects of proximity on network performance are).

We claim it depends on the level of proximity between agents whether their connection will lead to a higher level of innovative performance or not. The success of a network relation may be related to optimal levels of geographical proximity (Camagni, 1991), social proximity (Fleming et al., 2007; Uzzi, 1996), institutional proximity, organizational proximity (Grabher, 1993; Grabher and Stark, 1997) and cognitive proximity (Nooteboom, 2000). When thinking about an optimal level of geographical proximity, this does not mean determining an optimal geographical distance between two agents. Instead, one should think of a balance of local and non-local linkages. Similarly, the optimal social distance consists of a balance between embedded relationships within cliques and strategic 'structural hole' relationships among cliques. For institutional proximity, an optimal level consists of operating simultaneously in different institutional regimes, such as multinationals operating in different countries or high-tech labs cooperating with industry, government and academia. Concerning the optimal level of organizational proximity, loosely coupled networks that consist of weak ties between autonomous agents combine the advantages of organizational flexibility and coordination. The optimal level of cognitive proximity follows from the need to keep some cognitive distance (to stimulate new ideas through recombination) but also secure some cognitive proximity (to enable communication and effective knowledge transfer).

Besides looking for the optimal level of proximity on all dimensions, one can think of other solutions to the proximity paradox. The negative impact of excessive proximity in one dimension on innovative performance may be counteracted by lower levels of proximity in other dimensions. For instance, regions may confront the problem of regional lock-in by having a (related) variety of different technologies in the region (Frenken et al., 2007), or by having loosely coupled networks, as reflected in regional networks consisting of agents with weak ties (Grabher and Stark, 1997). In sum, optimal levels of proximity may enhance network performance, but the location of an optimum along one proximity dimension depends most likely on the location along other proximity dimensions at the same time.

Though the concept of optimal level of proximity balancing pros and cons has become well established (Boschma, 2005), to test these propositions empirically is not straightforward. There are two ways to go about this. First, classifying relationships into relations with high and low proximity, one can assess whether a mix of the two types of relationship leads organizations to perform better than organizations relying primarily on relations with low proximity or on relations with high proximity. This methodological strategy was followed by Uzzi (1996) to test the hypothesis of an optimal social proximity, and showed that a mixture of low and high proximity was best for firms. This has also been done in the case of geographical proximity: some have suggested a mixture of local and non-local linkages to be best for firms, and a combination of local buzz and global pipelines to be best for the long-term evolution of clusters (Bathelt et al., 2004). The second strategy is to classify all relations along a continuum and to assess the success of each particular relation separately. Then, by testing its effect and its quadratic effect, one can assess whether an optimal level of proximity exists (the linear effect should then be positive and the quadratic effect should be negative). For example, making use of patent data, Gilsing et al. (2007) assessed the effect of technological distance between firms in alliance networks in high-tech industries on the exploration innovative performance of firms. As expected, they found an inverse U-shaped function between technological distance and exploration.

4. Network dynamics

A key empirical insight from studies on networks, be it in the context of innovation and knowledge production or in other contexts, holds that networks have very pronounced structures (Newman, 2003). What we mean by structure is that the set of links between nodes in a network are very different from the properties of a random network, that is, the properties one obtains by randomly connecting nodes to create a network structure. Structured (or 'organized') networks require a true explanation, while random networks can simply be 'explained' stochastically.

Random networks are characterized by two important features. First, the degree distribution follows a normal distribution, where the degree of a node stands for the number of links of a node. Since a random network is constructed by assigning links between two randomly selected nodes, the degree of nodes will follow a normal distribution. Second, in a random network, there is no clustering: the probability of two nodes being linked is totally independent of whether these two nodes are linked indirectly via a third node. These two properties of random networks – normal degree distribution and absence of clustering – are never observed in social networks or inter-firm networks. Empirically, one typically observes that the degree distribution is skewed, with few nodes having a high degree and many nodes having a low degree. Apparently, some nodes are more 'popular' to link with than other nodes. And, one observes that clustering is a very significant phenomenon ('friends of friends are often friends with one another'). That is, many nodes participate in triangle relationships. Yet, some nodes do so much more than other nodes. The extent to which a node is clustered can be indicated by the number of triangles divided by the number of possible triangles.

At the level of single nodes, these observations lead to two questions: (1) How can one explain differences in the degree of nodes? and (2) How can one explain the clustering of nodes? Below, we discuss these features using the concepts of preferential attachment and closure, respectively. Then we propose an industry lifecycle perspective on network evolution and regional lock-in.

Preferential attachment

One key conceptual breakthrough in the study of dynamic networks has been the paper by Barabasi and Albert (1999). In this paper, the authors start from the observation that many networks are characterized by scale-free degree distributions where degree stands for node connectivity. They propose a simple growth model in which at each time step a new node is added and preferentially attaches itself to the node with the highest degree. More precisely, the probability that a new node attaches itself to an existing node is exactly proportional to the latter's degree. The specification of this mechanism reflects the benefits of linking to nodes with high degrees, as such 'hubs' provide new nodes with short pathways to many other nodes in the network. Assuming each node attaches to only one existing node, this mechanism leads to a power law distribution with an exponent equal to three.

In reality, most networks have degree distributions that are different from the pure Barabasi–Albert model. In particular, the degree of the best connected nodes is generally less than the model predicts. Indeed, the tendency of firms to connect to highly connected firms is found to be not as strong (Powell et al., 2005; Ter Wal, 2009). One explanation holds that firms are limited in the number of network relations they can meaningfully maintain. In the case of inter-firm networks, it is obvious there are limits to the number of partners a firm can maintain (Holme et al., 2004). This implies that well-connected nodes typically refuse proposals for networking and will select only the most beneficial partners (Giuliani 2007). A second reason why the degree distribution is less skewed than one would predict from preferential attachment is that proximity matters. This means that new nodes – even though attracted by the ones with highest connectivity – often connect to nodes with lower degrees if these are more proximate in any of the five dimensions we outlined before. Consider, for example, geographical proximity. A company may opt to collaborate locally to save on travel time and transportation costs, even though companies with the highest connectivity are located in other countries. The preferential attachment model can be easily adapted to incorporate this effect of proximity by assuming that the probability of a node linking to an existing node is not only dependent on the latter's degree but also on the geographical proximity between them (Guimerà and Amaral, 2004). The same reasoning holds for other forms of proximity. Depending on the benefits of proximity, such a constraint yields different network structures, ranging from very skewed degree distributions and low clustering as in the original Barabasi–Albert model when overcoming distance (in whatever dimension) is cheap, to networks with a normal degree distribution and a high clustering as in small worlds (Watts and Strogatz, 1998; Zhang et al., 2004) when overcoming distance is rather expensive, to an empty network where any relation is just too expensive to establish.

Closure

Another driving force of network formation is closure. In many instances, new network relations follow from existing relations as two actors are introduced to one another by a third actor with whom both already have a relation. The probability of the two actors forming a relation, who already relate to a common third, is expected to be much higher than the probability of two actors forming a relation who do not relate to a common third. The establishment of such triangle relationships is called 'closure' and such a closure mechanism will increase the degree of clustering (in a network sense) over time. The reason for closure to be common is twofold: (1) each actor can be informed by the common third about the properties of the other (what knowledge it possesses) and trustworthiness of the other, and (2) once the relationship is formed each actor has less incentive to behave opportunistically because of loss of reputation regarding the common third.

Note here that the role of social proximity in the formation of network links relates to the concept of closure. Viewing social proximity between two actors as the inverse of geodesic distance (network distance) in a network, closure simply means that if two actors have a social distance of two, they have a higher probability of getting connected. More generally, one expects the probability of a link to be formed to decrease with an increase in geodesic distance between two actors. Dynamically, this means that one expects the social proximity in networks to increase over time.

One hypothesis that has been analysed in a study by Ter Wal (2009) holds that closure in particularly relevant as a mechanism of network formation in exploitation contexts, while it is less important in exploration contexts. The reasoning underlying this hypothesis holds that closure is a way to find a new partner through an existing trusted partner, so that the collaboration with the new partner is embedded in the common relationship with the third actor. As a result, the partners in the new collaboration will have less incentive to behave opportunistically as they risk jeopardizing their relation with the third actor. Avoiding opportunistic behaviour is especially important in the exploitation phase of an industry, during which knowledge becomes more codified and is transformed in commercial products and services and, consequently, trust in partners is most important. And, logically, the formation of a new network relation is more likely, the more two actors already have partners in common. Studying the evolution of co-inventor networks in the German biotech industry in the period 1970 to 1995, Ter Wal (2009) found that closure was indeed a key factor driving network formation. As expected, closure also became more important over time as a driver of network formation with the biotech sector evolving from the exploration to the exploitation phase.

Proximity, industrial lifecycle and regional lock-in

As explained, the different forms of proximity likely influence the decisions of agents with whom to connect. As individuals and organizations prefer to establish relationships with similar type of individuals or organizations, network clustering will result as similar actors group together. When linking the proximity concept to the geography of innovation networks, major research challenges remain to be taken up. These have not yet (or hardly) been explored in the network literature, but are essential for the development of an evolutionary approach to the geography of innovation networks.

The main challenge is the study of the dynamics of network formation: how innovation networks of firms evolve in time and space, and what forms of proximity are important at what stage of the evolution of the network. The focus of attention is on the dynamics in the number of nodes and relations, and how the different forms of proximity impact on these network dynamics. It concerns both the study of: (1) the creation of relations by new firms entering the industry and by incumbent firms linking up with other nodes; and (2) the break-up of existing relations because of the exit of firms or because incumbent firms dissolve their relations with other nodes. Doing so, it covers the process of creative destruction proposed by Schumpeter, and applies that to the evolution of networks. Such an approach also accounts for the evolutionary concept of selection that basically takes place at two levels: (1) the impact of competition on firms leading to firm dynamics (i.e. the entry and exit of nodes); (2) the choice of linking or breaking with network partners (i.e. the formation and dissolution of ties) based on proximity. This means that firm dynamics are a basic input to understand the spatial formation of a network. As Klepper (2007) and others set out, spinoff dynamics is a crucial determinant of the location of industries, often leading to spatial clustering. In that respect, the emerging innovation network is most likely to cluster spatially as well, if not only because social relationships are established through the spinoff process between the parent organization and its offspring (i.e. the new spinoff companies).

Taking the industry lifecycle model as a point of departure, one can start to theorize about the network dynamics that follow. Studies have shown that after the creation of a new industry the number of firms first grows rapidly, then falls rapidly again (called a shake-out), and eventually stabilizes into an oligopolistic market structure dominated by a few persistent industry leaders (Klepper, 1997; Klepper and Simons, 1997). Furthermore, the spatial concentration of the industry tends to increase over time as successful parents create more, and more successful, spinoffs, which locate near their parents. After the shake-out, the firms that typically survive are indeed a few early entrants and their spinoffs. Apart from the famous case of spinoffs in Silicon Valley, examples can be drawn from the US and UK car industries (Boschma and Wenting, 2007; Klepper, 2007) as well as from the US tyre industry (Buenstorf and Klepper, 2005).

From the industry lifecycle pattern, we can derive propositions about the patterns of network evolution that are most likely to emerge (see e.g. Menzel and Fornahl, 2007; Ter Wal and Boschma, 2009b). First, as the knowledge base of an industry is progressively codified, the geographical distance of network relations is expected to increase over time (Menzel, 2008). This has indeed been observed in German inventor networks in the biotechnology sector (Ter Wal, 2009). Second, one can expect the probability of surviving a shake-out to be dependent on the degree of a firm in the inter-firm network. This means that the average degree of firms increases over time. Third, given the second proposition, the falling number of firms implies that the density over relations increases over time. Fourth, as spinoffs typically have a high degree of proximity with their parent firms in the cognitive, social and geographical dimensions, network relations between spinoffs and parents firms are much more likely than any other network relation type. The resulting geography of networks is, on the one hand, characterized by an increasing number of local links between spinoffs and parent firms in the same cluster. At the same time, one expects an increasing number of global links as a result of the increasing codification of knowledge. Thus, even though globalization of networks is expected to occur, the local density of network links is also expected to increase over time.

The industry lifecycle perspective can thus explain that the high density of network relations within clusters may become excessive as time passes by. As the number of firms falls over time, the remaining firms are typically embedded in strong social networks and interlocking corporate boards, which tend to resist structural change in the face of a crisis. Such resistance can be reinforced by increasing organizational proximity between firms through mutual financial participation between cluster firms as well as by higher levels of cognitive proximity between cluster firms, resulting from the long-lasting interactions in the past. According to Grabher (1993) and Hassink (2005), such structures typically explain the inabilities of old industrial regions to successfully renew themselves. The solution to such regional lock-in phenomena clearly lies in trying to re-organize network relations such that interactions can take place between actors that are less proximate in geographical and non-geographical dimensions. This could be accomplished by the formation of new ties that bridge unconnected networks (Burt, 2004; Glückler, 2007). These ideas call for further refinements and thorough empirical testing (Ter Wal and Boschma, 2009b).

5. Conclusion

We have made an attempt to sketch an evolutionary view of the geography of innovation networks by linking the literatures on proximity and network dynamics. To begin with, we argued that variety is a key feature of any economy, and knowledge accumulation at the firm level is its prime mover. In such an evolutionary framework of heterogeneous actors, the replication of knowledge between firms is considered troublesome unless there is some degree of proximity between actors on some dimensions: proximity is required on some (but not necessarily all) dimensions to make firms connected, and to enable interactive learning and innovation. Doing so, we have put the proximity concept into the heart of the theoretical and analytical framework of evolutionary economic geography.

Such a basic framework also enabled us to connect the proximity concept to the geography of networks. We made a distinction between five forms of proximity. Each relationship between two heterogeneous actors can be classified as being more or less proximate in all five dimensions. The dimensions, analytically defined, are orthogonal, even though many dimensions may often turn out to be correlated. A proximity framework suggests that actors that are proximate in some (if not all) dimensions are more likely to connect. This approach has led to new insights in the cluster literature, for instance. Giuliani (2007) has shown that knowledge networks between firms in a cluster are not pervasive (as suggested by the cluster literature) but tend to be rather selective, because these depend on the levels of cognitive proximity between cluster firms.

While a high degree of proximity is considered a prerequisite to make actors connected, we expect the effects of network relations on innovation to be rather ambiguous. Proximity between actors does not necessarily translate into higher innovative performance, because excess of proximity may be harmful for interactive learning. We referred to this as the proximity paradox. One should therefore make a distinction between the drivers of network formation on the one hand (in which the forms of proximity positively affect the establishment of networks), and the effects of a network on innovative performance on the other hand (in which it is uncertain what the effects of proximity on network performance are). Inspired by others, we expect that, for each dimension, an optimal level of distance exists, at which interactive learning and innovation are maximized.

We also introduced some propositions about network evolution and the changing role of proximity during the industry lifecycle. This has led us to conclude that the density of network relations in geographical clusters is likely to increase over time, despite the fact that codification of knowledge facilitates long-distance networking. A high local density of network relations may well lie at the root of the problems faced by industrial areas as a too strong proximity prevents the renewal of a region's industrial base.

We would like to finish with three research challenges that need to be taken up to build a true evolutionary perspective on the spatial evolution of innovation networks.

First, a dynamic network approach should assess the relative importance of the different forms of proximity as driving forces of network formation in space. This would not only shed light on the question of whether the different proximities are substitutes or complementarities, but also on the question of in which stage of the network formation some dimensions play a more prominent role. For example, if geographical proximity affects network formation, is this influence persistent over time? This concerns both the study of: (1) the creation of relations by new firms and by incumbent firms linking up with other nodes; and (2) the break-up of existing relations because of the exit of firms or because incumbent firms dissolve their relations with other nodes. Doing so, the study of network formation is not only about who connects with whom and why (being dependent on proximity), but it should also be grounded in firm dynamics (which concerns the formation and dissolution of nodes).

Second, a dynamic network approach should make explicit how the evolution of a network structure may be seen as a path dependence process, and how that may be tested. When the current structure of a network is affecting its future structure, network evolution becomes an endogenous process: the creation of a new tie is not only influenced by the structure of the network but it also causes changes in the network (Glückler, 2007; Kilduff and Tsai, 2003). Path dependence in network evolution is shown in the persistence of existing ties and the path-dependent formation of new ties. In the latter case, new ties replicate or reinforce the existing structure of the network (Gulati, 1999). The different proximities (besides preferential attachment) may induce path dependence in network evolution, and may cause retention in the network. These 'retention mechanisms' (Glückler, 2007) may take place at the local level because geographical proximity plays a role, both directly and indirectly (through its effect on the other proximities). If geographical proximity matters a lot in this respect, another crucial research question is under what conditions local network retention leads to regional lock-in, and how that may be broken apart.

Third, a dynamic network approach should also account for that fact that the evolution of network structures may, in turn, affect the degree of the different forms of proximity. The study on the dynamics of proximities during network formation is an (yet) unexplored but promising field of research: it would account for the effect of networks on the attributes of nodes in the network and thus their degree of proximity in their different dimensions over time, and how that might feedback on the structure of the network (Ter Wal and Boschma, 2009b). Moreover, a change in one proximity dimension can also have consequences for the other dimensions of proximity (Menzel, 2008).

In sum, we proposed an evolutionary perspective on the spatial evolution of network formation that is firmly grounded in a proximity framework. There are still many problems to be tackled before a dynamic proximity framework can be fruitfully applied to the spatial formation of networks. Having said that, we firmly believe it opens up a whole new research agenda that will contribute to a better understanding of the spatial evolution of innovation networks. In that respect, it may be a considered a crucial part of the further development of an evolutionary approach in economic geography (Boschma and Frenken, 2006).

References

Agrawal, A., I. Cockburn and J. McHale (2006), 'Gone but not forgotten: knowledge flows, labor mobility, and enduring social relationships', *Journal of Economic Geography*, **6**, 571–91.

Barabasi, A.L. and R. Albert (1999), 'Emergence of scaling in random networks', *Science*, **286** (5439), 509–12.
Bathelt, H., A. Malmberg and P. Maskell (2004), 'Clusters and knowledge: local buzz, global pipelines and the process of knowledge creation', *Progress in Human Geography*, **28** (1), 31–56.
Boschma, R.A. (2005), 'Proximity and innovation. A critical assessment', *Regional Studies*, **39** (1), 61–74.
Boschma, R.A. and K. Frenken (2006), 'Why is economic geography not an evolutionary science? Towards an evolutionary economic geography', *Journal of Economic Geography*, **6** (3), 273–302.
Boschma, R.A. and A.L.J. Ter Wal (2007), 'Knowledge networks and innovative performance in an industrial district: the case of a footwear district in the South of Italy', *Industry and Innovation*, **14** (2), 177–99.
Boschma, R.A. and R. Wenting (2007), 'The spatial evolution of the British automobile industry: does location matter?', *Industry and Corporate Change*, **16** (2), 213–38.
Breschi, S. and F. Lissoni (2003), 'Mobility and social networks: localised knowledge spillovers revisited', CESPRI Working Paper 142, Bocconi University, Milan.
Breschi, S. and F. Lissoni (2006), 'Mobility of inventors and the geography of knowledge spillovers. New evidence on US data', CESPRI Working Paper 184, Bocconi University, Milan.
Broekel, T. and A. Meder (2008), 'The bright and dark side of cooperation for regional innovation performance', Papers in Evolutionary Economic Geography, No. 08.11, Department of Economic Geography, Utrecht University.
Buenstorf, G. and S. Klepper (2005), 'Heritage and agglomeration: the Akron tire cluster revisited', Papers on Economics and Evolution 2005–08, Max Planck Institute of Economics, Evolutionary Economics Group.
Burt, R.S. (2004), 'Structural holes and good ideas', *American Journal of Sociology*, **110** (2), 349–99.
Camagni, R. (ed.) (1991), *Innovation Networks. Spatial Perspectives*, London/New York: Bellhaven Press.
Cohen, W.M. and D.A. Levinthal (1990), 'Absorptive capacity: a new perspective on learning and innovation', *Administrative Science Quarterly*, **35**, 128–52.
Edquist, C. and B. Johnson (1997), 'Institutions and organisations in systems of innovation', in C. Edquist (ed.), *Systems of Innovation. Technologies, Institutions and Organisations*, London, Pinter Publishers, pp. 41–63.
Etzkowitz, H. and L. Leydesdorff (2000), 'The dynamics of innovation: from national systems and "Mode 2" to a triple helix of university–industry–government relations', *Research Policy*, **29** (2), 109–23.
Fleming, L., S. Mingo and D. Chen (2007), 'Collaborative brokerage, generative creativity and creative success', *Administrative Science Quarterly*, **52** (3), 443–75.
Freeman, C. (1991), 'Networks of innovators: a synthesis of research issues', *Research Policy*, **20**, 499–514.
Frenken, K., F.G. van Oort and T. Verburg (2007), 'Related variety, unrelated variety and regional economic growth', *Regional Studies*, **41** (5), 685–97.
Gertler, M.S. (1995), '"Being there": proximity, organization, and culture in the development and adoption of advanced manufacturing technologies', *Economic Geography*, **71**, 1–26.
Gilsing, V., B. Nooteboom, W. Vanhaverbeke, G. Duysters and A. van den Oord (2007), 'Network embeddedness and the exploration of novel technologies. Technological distance, betweenness centrality and density', *Research Policy*, **37**, 1717–31.
Giuliani, E. (2007), 'The selective nature of knowledge networks in clusters: evidence from the wine industry', *Journal of Economic Geography*, **7** (2), 139–68.
Giuliani, E. and M. Bell (2005), 'The micro-determinants of meso-level learning and innovation: evidence from a Chilean wine cluster', *Research Policy*, **34** (1), 47–68.
Glückler, J. (2007), 'Economic geography and the evolution of networks', *Journal of Economic Geography*, **7** (5), 619–34.
Grabher, G. (1993), 'The weakness of strong ties – the lock-in of regional development in the Ruhr area', in G. Grabher (ed.), *The Embedded Firm*, London: Routledge, pp. 255–77.
Grabher, G. and D. Stark (1997), 'Organizing diversity: evolutionary theory, network analysis and postsocialism', *Regional Studies*, **31**, 533–44.
Granovetter, M. (1985), 'Economic action and social structure: the problem of embeddedness', *American Journal of Sociology*, **91**, 481–510.
Guimerà, R. and L.A.N. Amaral (2004), 'Modelling the world-wide airport network', *European Physical Journal B*, **38** (2), 381–5.
Gulati, R. (1999), 'Network location and learning. The influence of network resources and firm capabilities on alliance formation', *Strategic Management Journal*, **20**, 397–420.
Hagedoorn, J. (2002). 'Inter-firm R&D partnerships: an overview of major trends and patterns since 1960', *Research Policy*, **31** (4), 477–92.
Hall, P.A. and D. Soskice (eds) (2001), *Varieties of Capitalism. The Institutional Foundations of Comparative Advantage*, Oxford: Oxford University Press.
Hassink, R. (2005), 'How to unlock regional economies from path dependency? From learning region to learning cluster', *European Planning Studies*, **13** (4), 521–35.

Hoekman, J., K. Frenken and F. Van Oort (2008), 'Collaboration networks as carriers of knowledge spillovers: evidence from EU27 regions', CESPRI Working Paper 222, Bocconi University, Milan.
Hoekman, J., K. Frenken and F. Van Oort (2009), 'The geography of collaborative knowledge production in Europe', *Annals of Regional Science*, **43** (3), 721–38.
Hofstede, G. (1991), *Cultures and Organizations. Software of the Mind*, London: McGraw-Hill.
Holme, P., C.R. Edling and F. Liljeros (2004), 'Structure and time evolution of an Internet dating community', *Social Networks*, **26**, 155–74.
Kilduff, M. and W. Tsai (2003), *Social Networks and Organizations*, London: Sage Publications.
Klepper, S. (1997), 'Industry life-cycles', *Industrial and Corporate Change*, **6**, 145–82.
Klepper, S. (2007), 'Disagreements, spinoffs, and the evolution of Detroit as the capital of the U.S. automobile industry', *Management Science*, **53**, 616–31.
Klepper, S. and K.L. Simons (1997), 'Technological extinctions of industrial firms: an inquiry into their nature and causes', *Industrial and Corporate Change*, **6**, 379–460.
Knoben, J. (2007), *A Spider and the Stickiness of its Web. The Causes and Consequences of Spatial Firm Mobility*, unpublished PhD thesis, Tilburg: Tilburg University.
Lagendijk, A. and P. Oinas (eds), (2005), *Proximity, Distance and Diversity. Issues on Economic Interaction and Local Development*, Aldershot: Ashgate.
Maggioni, M.A., M. Nosvelli and T.E. Uberti (2007), 'Space vs. networks in the geography of innovation: a European analysis', *Papers in Regional Science*, **86** (3), 471–93.
Menzel, M.P. (2008), 'Dynamic proximities. Changing relations by creating and bridging distances', Papers in Evolutionary Economic Geography, No. 08.16, Department of Economic Geography, Utrecht University.
Menzel, M.P. and D. Fornahl (2007), 'Cluster life cycles. Dimensions and rationales of cluster development', Jena Economic Research Papers 2007–076.
Morrison, A. (2008), 'Gatekeepers of knowledge within industrial districts: who they are, how they interact', *Regional Studies*, **42**, 817–35.
Nelson, R.R. and S.G. Winter (1982), *An Evolutionary Theory of Economic Change*, Cambridge, MA: The Belknap Press of Harvard University Press.
Newman, M.E.J. (2003), 'The structure and function of complex networks', *SIAM Review*, **45**, 167–256.
Nooteboom, B. (1999), *Inter-Firm Alliances: Analysis and Design*, London and New York: Routledge.
Nooteboom, B. (2000), *Learning and Innovation in Organizations and Economies*, Oxford: Oxford University Press.
Nooteboom, B., W. Van Haverbeke, G. Duysters, V. Gilsing and A. van den Oord (2007), 'Optimal cognitive distance and absorptive capacity', *Research Policy*, **36** (7), 1016–34.
Ozman, M. (2009), 'Inter-firm networks and innovation: a survey of literature', *Economics of Innovation and New Technology*, **18** (1), 39–67.
Ponds, R., F.G. van Oort and K. Frenken (2007), 'The geographical and institutional proximity of research collaboration', *Papers in Regional Science*, **86**, 423–43.
Powell, W.W., K. Koput and L. Smith-Doerr (1996), 'Interorganizational collaboration and the locus of innovation: networks of learning in biotechnology', *Administrative Science Quarterly*, **41**, 116–45.
Powell, W.W., D.R. White, K.W. Koput and J. Owen-Smith (2005), 'Network dynamics and field evolution: the growth of interorganizational collaboration in the life sciences', *American Journal of Sociology*, **110** (4), 1132–205.
Rallet, A. (1993), 'Choix de proximité et processus d'innovation technologique', [Choice of proximity and technological innovation process], *Revue d'Economie Régionale et Urbaine*, numéro special 'Economie de Proximités', **3**, 365–86.
Rallet, A. and A. Torre (1999), 'Is geographical proximity necessary in the innovation networks in the era of global economy?', *Geojournal*, **49** (4), 373–80.
Singh, J. (2005), 'Collaborative networks as determinants of knowledge diffusion patterns', *Management Science*, **51** (5), 756–70.
Sorenson, O., J.W. Rivkin and L. Fleming (2006), 'Complexity, networks and knowledge flow', *Research Policy*, **35** (7), 994–1017.
Ter Wal, A.L.J. (2009), *The Spatial Dynamics of the Inventor Network in German Biotechnology. Geographical Proximity versus Triadic Closure*, Utrecht: Department of Economic Geography.
Ter Wal, A.L.J. and R.A. Boschma (2009a), 'Applying social network analysis in economic geography: framing some key analytical issues', *Annals of Regional Science*, **43** (3), 739–56.
Ter Wal, A.L.J. and R.A. Boschma (2009b), 'Co-evolution of firms, industries and networks in space', *Regional Studies*, DOI: 10.1080/00343400802662658, in press.
Torre, A. (2008), 'On the role played by temporary geographical proximity in knowledge transmission', *Regional Studies*, **42** (6), 869–89.
Torre, A. and A. Rallet (2005), 'Proximity and localization', *Regional Studies*, **39** (1), 47–60.

Uzzi, B. (1996), 'The sources and consequences of embeddedness for the economic performance of organizations: the network effect', *American Sociological Review*, **61**, 674–98.
Watts, D.J. and S.H. Strogatz (1998), 'Collective dynamics of "small-world" networks', *Nature*, **393**, 440–42.
Weterings, A. (2005), *Do Firms Benefit from Spatial Proximity? Testing the Relation between Spatial Proximity and the Performance of Small Software Firms in the Netherlands*, Netherlands Geographical Studies, No. 336, KNAG: Utrecht.
Williamson, O.E. (1985), *The Economic Institutions of Capitalism*, New York: The Free Press.
Zhang, Z.-Z., L.-L. Rong and F. Comellas (2006), 'Evolving small world networks with geographical attachment preference', *Journal of Physics A: Mathematical and General*, **39**, 3253–61.

PART 2

FIRM DYNAMICS, INDUSTRIAL DYNAMICS AND SPATIAL CLUSTERING

6 Entrepreneurship, evolution and geography

Erik Stam

1. Introduction

Entrepreneurship is a fundamental driver of economic evolution. It is also a distinctly spatially uneven process, and thus an important explanation of the uneven economic development of regions and nations. Not surprisingly, entrepreneurship is a key element of evolutionary economics (Grebel, 2004; Grebel et al., 2003; Metcalfe, 2004; Schumpeter, 1934; Witt, 1998) and has been recognized as an important element in explaining (regional) economic development (Acs and Armington, 2004; Audretsch et al., 2006; Fritsch, 2008). This means that the explanation of regional variations in entrepreneurship has also become an important issue. Even more so because there are pronounced differences within and between nations in rates of entrepreneurship and in their determinants (Bosma and Schutjens, 2008), and these differences tend to be persistent over time, reflecting path dependence in industry structure (Brenner and Fornahl, 2008), institutions (Casper, 2007) and culture (Saxenian, 1994) that vary widely across regions and countries, but are relatively inert over time. Introducing entrepreneurship into evolutionary economic geography means that the traditional focus on firms is complemented with a focus on individuals.

This chapter is an enquiry into the role of entrepreneurship in evolutionary economic geography. The focus is on how and why entrepreneurship is a distinctly spatially uneven process. We start with a discussion on the role of entrepreneurship in the theory of economic evolution. Next, we review the empirical literature on the geography of entrepreneurship. The chapter concludes with a discussion of a future agenda for the study of entrepreneurship within evolutionary economic geography.

2. Entrepreneurship and economic evolution

'Newcomers' to the economy have an important role to play in the evolution of economic systems. According to Schumpeter (1942; p. 83):

> The fundamental impulse that sets and keeps the capitalist engine in motion comes from the newcomers' goods, the new methods of production or transportation, the new markets, the new forms of industrial organisation that capitalist enterprise creates. . . . [This is a] process of industrial mutation – if I may use that biological term – that incessantly revolutionises the economic structure *from within*, incessantly destroying the old one, incessantly creating a new one. This process of Creative Destruction is the essential fact about capitalism.

By creating new variations (products, processes, business models) in the economy, these innovative new firms compete with incumbent firms, which force the latter to improve or change their production, sanctioned by liquidation if this is not done successfully (Schumpeter, 1934, 1942). The creation of this variation is unevenly distributed over space. Although relatively inert, this spatial distribution of variety creation itself changes over time. These new variations are thus created somewhere, but are not diffused automatically to all places and applications in which they might be of value. A less heroic, but perhaps not less important role is played by entrepreneurs in this diffusion of new variations: they fill the gaps in the market (Kirzner, 1973, 1997). Introducing existing products and practices to new contexts – via processes of generalization, differentiation, or reciprocation (see Nooteboom, 2008) – can be a truly entrepreneurial effort that might even lead to radical innovations. Variety creation and diffusion are two important roles played by the entrepreneur in economic dynamics. For example, the formation of new technology based firms might serve the purpose of creating new – technology intensive – products or of diffusing (the use of) new technologies in society.

Variation and diffusion also feed each other (see Nooteboom, 2008): the pursuit of entrepreneurial opportunities feeds further opportunities (Holcombe, 2007). First, any change by one entrepreneur alters the economic environment and provides opportunities for additional adjustments by other entrepreneurs. Second, entrepreneurial activity is likely to create wealth and in that way increases the extent of the market. Third, the creation of market niches that did not previously exist provides opportunities for new entrepreneurs to enter and expand this market niche. Entrepreneurial opportunities come into being because of prior acts of entrepreneurship (see Metcalfe's, 2002, 'growth of knowledge'): 'Bill Gates could not have made his fortune had not Steve Jobs seen the opportunity to build and sell computers, and Steve Jobs could not have built a personal computer had not Gordon Moore invented the microprocessor' (Holcombe, 2007; p. 61).

Next to variation and diffusion, selection plays an important role in entrepreneurship,[1] reflected in the fact that most new firms do not survive for a long time, and that even a smaller portion (often less then one out of ten start-ups) grows to some extent (Reynolds and White, 1997; Stam et al., 2008). Selection is generated by the decisions of external resource holders to allocate their resources among these firms (Aldrich, 1999; Baum and Silverman, 2004). New firm formation is affected by different selection environments. Most directly there is competition in product-markets: a lack of competition might indicate an opportunity (a gap to be filled) and a constraint (with too high entry barriers). Fierce competition forces firms to produce and sell efficiently, in order to survive. For new firms that need to reach a substantial size, selection in the capital and labour market is also important. They need to attract finance and human resources in competition with other organizations that need these resources in the face of limited supply. Competition is often a very local process: more distant firms are less likely to compete for the same pool of human resources or product-markets than firms in proximity (Baum and Mezias, 1992; Cattani et al., 2003; Sorenson and Audia, 2000).

Historically, the literature has often explained entrepreneurship as either the product of environments (like provision of venture capital, growing demand) or of personal attributes (like risk-taking propensity, need for achievement). Individuals are heterogeneously endowed with skills, knowledge, attitudes and preferences (values) that drive their motives and behaviour (McFadden, 2001; Simon, 1957). Environments are heterogeneously endowed with knowledge, institutions, resources and demand for products. The entrepreneurial process depends on entrepreneurial opportunities in the environment and enterprising individuals that identify and exploit these opportunities. When individuals identify an opportunity, they do not react automatically by establishing a new firm (assuming that they have the intention to start one): new firms are created with a sequence of processes like creating a legal entity, product development, and financing (Carter et al., 1996). Given that prospects of employment, education, and other circumstances differ across individuals, the population is heterogeneous with different individuals facing different opportunity costs when acting to exploit an opportunity they have recognized. Entrepreneurship is the result of the interaction between individual attributes and the surrounding environment. This means that entrepreneurs are neither the lonely heroes that change the economy on their own, nor are they determined by their environment: just like any other individual they most often reproduce their structural conditions, but they are entrepreneurial because they also transform these structures. The latter echoes the Schumpeterian view of entrepreneurs as the executors of transformative new combinations, and involves the pursuit of entrepreneurial opportunities, defined as 'ideas, beliefs and actions that enable the creation of future goods and services in the absence of current markets for them' (Sarasvathy et al., 2003, p. 142).[2] A working definition of entrepreneurship in line with this view is 'the introduction of new economic activity by an individual that leads to change in the marketplace' (see Stam, 2008). This new activity can be perceived in reality as a new good or service that is produced by or for the entrepreneur, and that is valued by consumers who pay a price for its property rights. This excludes non-market activities (i.e. no price mechanism and property rights involved) and mere changes of contract (i.e. no new economic activities involved). Let us give three examples of what entrepreneurship is not according to this definition (and separating it from much everyday usage of the term). First, the shift from employment into self-employment by an individual does involve a change in the marketplace, but not an introduction of new economic activity. Second, the creation and introduction of a new product in a concerted effort by a large corporation that involves exchangeable individuals also does not count as entrepreneurship. Third, the creation and execution of a new terrorist strategy in which airplanes are used as missiles (the 11 September 2001 attack), involves new economic activities by a distinct group of individuals who might not be interchangeable, but does not lead to an exchange of property rights.

From a theoretical perspective, an inquiry into the role of entrepreneurship in evolutionary economic geography builds on insights from evolutionary economics, cognitive theories of innovation, social network approaches, and organizational ecology. These fields reveal large overlaps in the processes they study.[3]

Within evolutionary economics (individual and collective) learning processes, inheritance of routines and feedback effects play an important role. In evolutionary economics the variance in the performance of firms is explained by heterogeneity in routines (Hodgson and Knudsen, 2004; Nelson and Winter, 1982).[4] Routines can be understood as organizational skills, which cannot be reduced to the sum of individual skills, that is, they are a collective property.[5] However, it is still unclear what the role of individual-level skills and knowledge (an individual property) is in relation to organizational routines (a collective property). We will get back to this issue in the final section of this chapter. The replication of routines takes place between firms (as carriers of routines) through various mechanisms, of which one is the creation of a firm by an employee (Klepper, 2002, 2007; Klepper and Sleeper, 2005) through which routines (and the knowledge embedded in them) are transferred from the parent to the newly created firm. Next to the emphasis on the replication of routines, evolutionary economics' conceptualization of economic evolution as the emergence and dissemination of novelty (Witt, 2003) moves the entrepreneur as a creator and disseminator of novelty in the economy centre stage. The emergence of novelty creates variety in any evolving system, but the generation of novelty requires heterogeneous elements as inputs to the recombination process underlying it (Witt, 2005). This brings us to cognitive theories.

Cognitive theories of innovation emphasize that innovation is a product of interaction between actors that have sufficiently different knowledge in order to make transformative (Schumpeterian) new combinations, but are still sufficiently proximate in a cognitive sense in order to be able to communicate at all (Nooteboom, 2000). On the micro level these innovations are most likely to be realized by spin-off firms pursuing opportunities that are based on the existing knowledge base of the parent firm, but sufficiently different to exploit it outside the parent organization. Empirical studies have shown that industries like instruments manufacturers (Audia et al., 2006) and automobiles (Boschma and Wenting, 2007; Carroll et al., 1996; Klepper, 2002) have emerged in this way: the successful early entrants in the automobile industry came from related bicycle producers, carriage builders, and engine manufacturers, while the successful early entrants in the instrumentation industry came for example from machine, defence, and chemicals industries (Audia et al., 2006).

Organizational ecology studies populations of organizations, focusing on how they change over time, especially through demographic processes of selective replacement – organizational founding and mortality (Carroll and Khessina, 2005). The evolutionary triad of variation, heredity and selection is central in the organizational ecology approach. Organizational foundings are predicted with notions like density dependence, structural inertia, niche width, and resource partitioning (see Carroll and Hannan, 2000). Organizational density is driven by organizational foundings and affects competition and legitimacy of a particular organizational form. Organizational inertia and imprinting are important mechanisms of retention. In this field new firms are often analysed as organizational products (Audia et al., 2006; Audia and Rider, 2006; Freeman, 1986).

Finally, network studies emphasize the role of information acquisition and resource mobilization via social networks in the behaviour of individuals and groups. Key issues related to entrepreneurship are processes of opportunity identification and resource mobilization (Sorenson, 2003; Stuart and Sorenson, 2007). These literatures all take into account the role of entrepreneurship in creating something new, which is somehow related to the past, and is affected by and affects its context.

3. Regional conditions of entrepreneurship

Entrepreneurs are hardly lone individuals who rely primarily on their extraordinary efforts and talents to overcome the difficulties inherent in the formation of a new firm. The process of starting a new firm is eminently social, as information and resources are to a large extent acquired via the personal networks of the (nascent) entrepreneur. For nascent entrepreneurs the focal choice is what kind of firm to start given their location, not so much choosing a location for a given firm (Stam, 2007). The social ties of the potential entrepreneurs are likely to be localized, and induce entrepreneurs to start their firm in close proximity to their homes and to their current employers (Cooper and Folta, 2000; Parwada, 2008; Sorenson, 2003; Stam, 2007). It is a stylized fact that entrepreneurs start their firm in the region where they live and/or work. The fraction of entrepreneurs working in the region where they were born is significantly higher than the corresponding fraction for dependent workers (Michelacci and Silva, 2007). A study of Portuguese manufacturing firms found that entrepreneurs were willing to accept labour costs three times higher than in alternative locations to locate the new business in their current region (Figueiredo et al., 2002).

There are several reasons for the locational inertia of entrepreneurs. First, they can utilize their existing (local) network to seek partners, employees, suppliers, customers, advisors and investors (Michelacci and Silva, 2007; Zander, 2004). This decreases search costs, but it also permits them to build on credibility and trust developed in past relationships. The behavioural matrix of Pred (1967) is relevant here, as locational inertia can also be explained by imperfect information about alternative locations and/or limited cognitive abilities to process all information available (see Simon's, 1957, bounded rationality). Second, more normative motivations might be at play here, as some relationships involve more than rational instrumental motivations, and continuing these relationships might only be possible when the entrepreneur stays within the region. Dahl and Sorenson (2009) conclude in their empirical study of Danish entrepreneurs that entrepreneurs appear to value proximity to family and friends not for the help that those connections might offer to their ventures but for emotional reasons. Third, they can start on a part-time basis (often being home-based) and delay full-time commitment until the venture seems sufficiently promising (part-time entry as a real option strategy: see Wennberg et al., 2007). Third, a spouse can keep a job so that income continues to flow to the family; other aspects of a founder's life can remain the same (Hanson, 2003). The full energies of the entrepreneur can then be devoted to start-up.

Earlier in this chapter it has been said that entrepreneurship is the result of the interaction between individual attributes and the surrounding environment. For explaining the spatial distributions of entrepreneurship, one should thus look at spatial aspects of this interaction. In geographical terms we mainly refer to characteristics of particular places and spatial distance between particular actors. We can start the explanation with the availability of (potential) entrepreneurs in particular places. Several perspectives are useful here: the nature and number of organizations in a region, the regional culture, and the labour market structure in a region. Key elements are the resources, abilities and preferences of individuals. The key question is why in a given (opportunity) environment some individuals are more likely to start a firm than in another environment: for example because of their willingness to incur risk, preference for autonomy and self-direction, specific human capital and experience. One important underlying factor can be found in generational effects: having an entrepreneurial family background strengthens the probability of entering self-employment. Intergenerational transmission of self-employment is an explanation for spatial differences in self-employment (Niittykangas and Tervo, 2005; Vaillant and Lafuente, 2007).

Another starting point represents the opportunities for entrepreneurship. From this point of view, individuals in particular environments are more likely to be entrepreneurs because the availability of opportunities encourages their exploitation by starting a firm. The sources of opportunities can be manifold: for example a growing purchasing power in the region, technological change, regulatory change.

Historical processes produce uneven spatial economic patterns, of both the characteristics of individuals and the 'availability' of opportunities, that condition but do not determine economic behaviour (Boschma and Frenken, 2006), of which entrepreneurship is a special class. In the following sections we review the empirical literature that relates to entrepreneurship, evolution and geography.

Entrepreneurship as an organizational product

Although some individuals become successful entrepreneurs without related prior experience, they are the exception rather than the rule. Entrepreneurs are often organizational products, that is, they spin off a firm from their previous employer (Audia and Rider, 2005, 2006; Klepper, 2001). Many entrepreneurs are characterized by 'sectoral inertia', that is, they start their firm in an industry with which they were already familiar (Cross, 1981; Johnson and Cathcart, 1979; Lloyd and Mason, 1984; Storey, 1982; Vivarelli, 1991). Far from the universal choice, entrepreneurial action is relatively constrained: instead of looking around to seek the most profitable opportunity, the potential entrepreneur concentrates his or her attention on a familiar sector (Vivarelli, 1991). A person working in an industry is more likely to identify a market gap than a person without any industry experience (O'Farrell and Crouchley, 1984), irrespective of the degree of industry competition and growth prospects (Storey, 1982).

This prior experience (Shane, 2000) and personal networks are likely to be acquired during the entrepreneur's career in existing organizations (Agarwal et al., 2004; Gompers et al., 2005; Klepper, 2001). This explains why the nature and number of organizations in a region are important determinants of entrepreneurship in a region.

Empirical research has shown that regions dominated by small and/or young firms have relatively high new firm formation rates (Audretsch and Fritsch, 1994; Johnson and Cathcart, 1979; Mason, 1991; Mueller, 2006; Reynolds et al., 1994; Sørensen, 2007). This stylized fact may be caused by several mechanisms: experiential learning, vicarious learning, competition, and entry barriers. The latter three mechanisms will be discussed in the section on industry structure. In this section we discuss experiential learning as well as vicarious learning (peer effects). The greater the proportion of an industry's labour force with direct experience of working in smaller firms, the more widespread the propensity for self-employment and hence the greater the propensity to start a new firm. Industries dominated by large plants would be expected to perform poorly as incubators of new business founders (Gudgin, 1978, pp. 211–12; Johnson and Cathcart, 1979). Young organizations especially that were once venture capital-backed, that focused on one segment, and whose growth slowed, have high 'entrepreneurial spawning' levels (Gompers et al., 2005). Next to direct, experiential learning (learning how to set up and grow a business) peer effects are also important here. A study by Nanda and Sørensen (2008) showed that an individual is more likely to become an entrepreneur if his or her co-workers have been entrepreneurs before. They argue that peers matter in two ways for entrepreneurship: by structuring co-workers' access to information and resources that help identify entrepreneurial opportunities, and by influencing co-workers' perceptions about entrepreneurship as a career choice.

Klepper and colleagues (Klepper, 2001, 2002, 2006, 2007; Klepper and Simons, 2000; Klepper and Sleeper, 2005; and Buenstorf and Klepper, 2009; Agarwal et al., 2004; Buenstorf and Fornahl, 2009; see also Helfat and Lieberman, 2002; and Dahl et al., Chapter 9, this volume) have constructed a model in which spin-offs exploit knowledge from their parents. Firms are assumed to differ in terms of their initial competence at time of entry, which shapes long-term performance. This competence is acquired from firms in related industries and prior entrants in the same industry.

The stock of incumbents and firms in related industries in a region determines the entry rate and post-entry performance of firms in this particular industry. The differences in entry by region are not necessarily determined by differential distance to buyers and suppliers, thick labour markets, or to spill-overs between firms (more likely to be within firms, with employees spinning out afterwards). They explain the market conditions conducive to spin-offs, the types of firm that spawn spin-offs, and the relationship of spin-offs to their parents. The model is tested using detailed data on entrants in industries like automobiles (Boschma and Wenting, 2007; Carroll et al., 1996; Klepper, 2002, 2007), tyres (Buenstorf and Klepper, 2009), television receivers (Klepper and Simons, 2000), lasers (Klepper and Sleeper, 2005; Buenstorf, 2007), semiconductors (Moore and Davis, 2004), law firms (Phillips, 2002), accounting (Wezel et al., 2006), footwear (Sorenson and Audia 2000), fashion design (Wenting, 2008), and software (Buenstorf and Fornahl, 2009). Their findings support the basic premise of the model that spin-offs inherit knowledge from their parents that shapes their nature at birth and their survival chances.

Industry structure

New firm formation across regions can be explained by differences in the regional composition of industries and by differences in one particular industry in specific regions. The latter would indicate that there are context-specific differences affecting entrepreneurship rates, while the former would indicate that the explanation should largely be sought in the specific industry structure of the region. The industry structure of a region affects the overall new firm formation rates in a region, as industries differ in their degree of contestability (entry barriers) and the extent to which entrepreneurial opportunities emerge (e.g. many in business services and few in mining).

Sometimes both the industry structure and the regional context are favourable for new firm formation in a region; this can for example be found in the south-east of the UK, which has both a favourable industry mix (especially construction, service, and finance and related sectors) and favourable local conditions. In contrast, regions like Northern Ireland, Scotland and Wales suffer from a combination of both an unfavourable industry mix and unfavourable local conditions for new firm formation. Often the industry mix component dominates the local conditions component in statistical analyses of determinants of regional firm formation rates (Fotopoulos and Spence, 2001).[6]

Several mechanisms related to the industry structure can be at work in a region, to produce something that academics and policymakers like to call a cluster (Martin and Sunley, 2003). Two important concepts connected to clusters are localization economies and related variety. The first concept has a long history in the academic literature (Chinitz, 1961; Hoover, 1948; Malmberg and Maskell, 2002; Marshall, 1890), while the second one is only recently recognized in evolutionary economics (Frenken et al., 2007) and organizational ecology (Audia et al., 2006).

Localization economies involve agglomeration economies resulting from the concentration of the same or similar activities: for example, benefits resulting from the local access to a specialized work force or the specialized reputation of a locality, while related variety emphasizes the positive effects (on entrepreneurship, innovation) of the co-presence of different but related industries or organizational populations. Regions that have a concentration of organizations of a certain kind (in a specific industry) tend to generate a relatively large number of new organizations of that same kind. This pattern has been shown for industries in general (Armington and Acs, 2002), and for specific industries like footwear (Sorenson and Audia, 2000), accounting (Cattani et al., 2003), biotechnology (Stuart and Sorenson, 2003), computer workstations (Sorenson, 2005), motorcycles (Wezel, 2005).

There are several reasons for why industrial clusters foster entrepreneurship (Audia and Rider, 2005; Rocha, 2004). Clusters provide established relationships and better information about opportunities. They open up niches of specialization because of the low degree of vertical integration. Clusters foster a competitive climate and strong rivalry among firms that impose pressure to innovate because of the presence of close competitors. They provide role models with the presence of other firms that have 'made it' (see Fornahl, 2003; Vaillant and Lafuente, 2007), and a cultural environment where establishing one's own business is normal and failure is not a social stigma. Clusters provide access to physical, financial, and commercial infrastructure; easing the spin-offs of new companies from existing ones. Especially because potential entrants will know how the local industry functions and have the technical skills to operate in it.

Industry localization may have negative effects on new firm formation: increased concentration and vertical integration raise entry barriers (Beesley and Hamilton, 1994). The shift in the direction of the effect of localization on new firm formation might be explained by the life-cycle stage of the industry: in the early stages geographical concentration has positive effects (or is even driven by new firm formation; see Feldman, 2001; Feldman and Francis, 2003; Feldman et al., 2005), while in late stages (stagnant or even declining markets, and increased relevance of scale economies) the negative effects dominate. Spatial concentrations of activities in mature industries might still have high new firm formation rates (still cognitive effects and low barriers to entry), but high levels of competition (and decreasing demand) lower the performance of these entrants (Sorenson and Audia, 2000). This means that there is still industry localization, but there are no localization economies any more.

Urbanization

Urban areas have important advantages for entrepreneurship. Population density has been found to positively affect entrepreneurship (Reynolds et al., 1994; Wagner and Sternberg, 2004). Some authors have argued that this positive effect of population density (most emphasized in big cities) might be a temporary phenomenon: the resurgence of big cities in the 1990s is connected both to a reduction in the negative social interactions (e.g. crime) and to an increase in positive social interaction (Glaeser and Gottlieb, 2006). However, urbanization and its most straightforward indicators, population size and density, cover many mechanisms, which might have different weights and values in different contexts. This is for example reflected in the large heterogeneity in entrepreneurship rates in world cities (Acs et al., 2008). Advantages stemming from high population density include the relative ease of access to customers as well as the inputs required (capital, labour, suppliers) to produce the goods or services. The classical 'incubation hypothesis' in urban economics states that persons aspiring to go into production on a small scale have found themselves less obviously barred by a high cost structure at the centre of the urban area than at the periphery (Chinitz, 1961; Dumais et al., 1997; Hoover and Vernon, 1959). In addition, cities provide contexts in which serendipitous meetings are more likely to occur than in less densely populated areas (Jacobs, 1969); these serendipitous meetings increase the likelihood of new opportunities and collaborations that might trigger the emergence of a new firm. Urban density also improves the likelihood of getting into contact with more skilled individuals in the same or related knowledge domains: learning from these more skilled peers stimulates human capital accumulation in urban environments (Glaeser, 1999) and might lead to the creation and recognition of better quality entrepreneurial opportunities. This human capital effect on opportunities is strengthened by the relatively high concentration of universities and research centres in urban areas, which produce new scientific and technological knowledge that has also been recognized as an important source of entrepreneurial opportunities (Audretsch et al., 2006; see also the section on 'growth in knowledge'). The risk of starting a business in urban areas is also relatively low because of the abundant employment opportunities, which function as an occupational buffer for the entrepreneur when the firm fails.

Urbanized areas are often concentrations of educated individuals with business experience in their early and middle adult years,[7] and in that way they are a source of entrepreneurs (Glaeser, 2007). Urban areas have important advantages for the demand for entrepreneurship (especially in retailing), as they contain demand for a rich variety of services and consumer goods (see Glaeser, 2007; Glaeser et al., 2000). Urbanization positively affects diversification of consumer demand. This latter phenomenon is central in flexible specialization theory (Piore and Sabel, 1984), which explains such trends in terms of the breakup of the mass market for standardized goods and services and the consequent emergence of a variety of smaller niche markets capable of exploitation by new or small businesses. This diversification is, next to urbanization, also directly driven by growth in overall demand.

Culture

Culture is important in the explanation of spatial variation in entrepreneurship via its effect on the attitude and values that people acquire. Social psychologists have claimed that an individual's attitudes and traits are not inherited but are developed in interaction with the social environment. Perceptions about the desirability of becoming an entrepreneur are formed and revised given the set of information available to each person (Lafuente and Salas, 1989; Saxenian, 1990). Culture is a property of groups, and it seems that especially national (Uhlaner and Thurik, 2007), and to a lesser degree regional cultures (Davidsson and Wiklund, 1997) have significant effects on new firm formation. These cultures can change over time, but they tend to be very persistent (Beugelsdijk, 2007).

The existence of a number of entrepreneurs in a region also legitimizes the activities of nascent entrepreneurs (Kristensen, 1994). Differences in culture in that way affect the level to which people aspire and think about becoming an entrepreneur, which is an important phase in the process of starting a firm. One example of this is the fear of failure that might deter people from starting a new firm (Arenius and Minniti, 2005; Vaillant and Lafuente, 2007). Cognitive theories have proposed that individuals acquire information and skills by observation of (entrepreneurial) activities by others, that might trigger and enable their choice of an entrepreneurial career (Forbes, 1999; Minniti, 2005; Zander, 2004).[8]

Distinct local cultures can have effects on different types of entrepreneurship. For example a 'self-employment' and a 'career' life-mode have different positive effects on entrepreneurship, while a 'wage-earner' life-mode has a clear negative effect on entrepreneurship (Illeris, 1986). In the self-employment life-mode the dominant job-related motivation is to own the means of production and control the production process. This culture is carried over from generation to generation, and is most frequently found in rural areas characterized by independent and self-reliant small-scale farmers, artisans and small business owners. It is rare in areas dominated by large-scale operations. The dominant value in the wage-earner life-mode is the sale of one's labour at the highest possible price in order to maximize the utility of one's leisure time. Such individuals are unlikely to set up new businesses, except when they are 'forced' by unemployment. This life-mode is likely to be most common in regions characterized by a narrow industrial base and dominated by large externally owned firms. Finally, the dominant value of individuals with a career life-mode is the advancement of their career. They are likely to be well-educated and working in large private or public sector organizations. They will start their own businesses if this becomes the best way in which to benefit from their skills, knowledge and expertise. These businesses are often technologically advanced, innovative and with good marketing capabilities. Career mode entrepreneurs are often concentrated in large metropolitan areas and smaller attractive cities (Savage et al., 1988). This career life-mode resembles the life-mode of the so-called creative class to a large extent. The creative class consists of individuals with relatively high levels of creativity in their work. A spatial concentration of creative class has been shown to positively affect new firm formation rates (Lee et al., 2004; Marlet and Van Woerkens, 2007; Van Aalst et al., 2006). Indirectly, amenities in regions ('quality-of-place': Florida, 2002) affect new firm formation rates, by attracting the creative class.

Growth in knowledge

New knowledge created at universities and research centres generates opportunities for entrepreneurship, especially in high-tech industries. Often these organizations are not able to fully recognize and appropriate the ensuing opportunities to commercialize that knowledge. Knowledge workers in these organizations respond to opportunities generated by new knowledge by starting a new firm, and in this way appropriate the expected value of their endowment of knowledge (Acs et al., 2005; Audretsch et al., 2006; Feldman, 2001; Kirchhoff et al., 2007; Zucker et al., 1998).[9] Geographic proximity to these sources of new knowledge is an asset, if not a prerequisite, to entrepreneurial firms in accessing and absorbing spillovers from universities and research centres (Audretsch and Feldman, 1996; Audretsch and Lehmann, 2005a, 2005b; Audretsch et al., 2005; Audretsch and Stephan, 1996). The most relevant spatial scale where these spillovers take place is not clear, as these knowledge spillovers are said to extend up to approximately 10 km (Baldwin et al., 2008), 50 miles (Anselin et al., 1997), 75 miles (Varga, 1998), 145 miles (Woodward et al., 2006), or even 300 km from their source of origin (Botazzi and Peri, 2003). The temporal scale on which these processes take place might also range from a few months to several decades: major breakthrough inventions like the discovery of the DNA were followed by substantial entrepreneurial activity only decades later (and often in other places than the context of invention, Cambridge, UK, in this instance). Another mechanism, the creation of embodied knowledge, via education and learning by doing, also takes multiple years, and possibly at multiple locations.

The degree to which technological change promotes new firm formation (in high-tech industries) depends on the institutional environment. The institutional setting affects the nature of technical labour markets, venture capital markets, and the structure of buyer–supplier ties that are highly relevant for the incentive constraints and appropriability constraints acting on incumbent and start-up firms respectively (Casper, 2007; Chesbrough, 1999). For example institutions enabling a fluid labour market, a well developed venture capital market, and loose buyer–supplier ties allow new firms to rapidly assemble and deploy experienced engineering talent, and move quickly to commercialize advanced technology. Cross-country research has shown that this latter situation can be found in the US, in contrast to Japan (Chesbrough, 1999), and cross-regional research found for example that Silicon Valley was much more conducive to new technology-based firms than Route 128 (Massachusetts) for similar reasons (Kenney and Von Burg, 1999; Saxenian, 1994). The institutional environment also affects the opportunity costs involved in leaving a (relatively secure) job at a university or research centre for self-employment (see Feldman, 2001). The institutional environment thus acts as a mediating factor between investments in the knowledge base of a society and the knowledge spillovers exploited by entrepreneurs.

Financial capital

Liquidity constraints are an important factor disabling entrepreneurs to realize their business opportunities (Evans and Jovanovic, 1989; Holtz-Eakin et al., 1994). This is especially relevant for large new firms that require relatively large-scale investments for their initial activities. Small-scale firms can often be financed with bank loans or the support of the entrepreneur's family and friends. The entrepreneur's own housing is shown to be the single most important source of collateral for bank loans (Black et al., 1996). Indirectly, fluctuations in the local housing market could thus affect the availability of financial capital for new firms. New firms that require large-scale investments are more likely to enter the venture capital market. Providers of venture capital provide not only financial capital, but also knowledge of developing a business (in a particular industry). The provision of financial capital in general is more likely to be bounded to national scales, while the provision of venture capital is often constrained to regions (Gibbs, 1991; Zook, 2002). The supply of venture capital is not distributed evenly across regions. For example the venture capital market in the USA is highly concentrated (both in supply and investments) in the east and west coasts of the country (Powell et al., 2002; Sorenson and Stuart, 2001), and in the UK it is highly concentrated in the south-east, in and around London (Mason and Harrison, 1999, 2002; Martin et al., 2005). This uneven regional distribution has also been found in other countries (Martin et al., 2002). Venture capital markets are a relatively recent phenomenon and often co-evolve with other investment intensive industries in particular regions (see Braunerhjelm and Feldman, 2006). The uneven regional distribution of venture capital means that in regions far away from these centres entrepreneurs might be discouraged from starting capital-intensive firms. The assumption is that spatial proximity may be necessary for the formation of a venture capital relationship and that it makes monitoring of investments easier. Face-to-face contacts between the entrepreneur and the venture capital provider are necessary to identify the value of the new business and the involvement in business affairs by the venture capital provider. These contacts are hard to initiate and sustain over a large distance (Mason and Harrison, 2002; Sorenson and Stuart, 2001; Stuart and Sorenson, 2003). Recent evidence shows that most of these contacts cannot easily be maintained over a longer distance via telecommunication: this can be used as a complement to face-to-face contacts, not as a substitute (Fritsch and Schilder, 2008).

4. Summary and research agenda

In the prior sections we have reviewed the empirical literature that relates to entrepreneurship, evolution and geography. The industry structure of a region was revealed to have consistent effects on the prevalence of entrepreneurship in a region. This covers both the nature of the activities as well as the characteristics of the organizations (size and age). Regional and national culture conditions the preference for entrepreneurship. Another important factor for entrepreneurship is the expanding knowledge base of a region, which increases the number of technological inventions to be commercialized by new firms. Urbanization is also likely to affect entrepreneurship, as it generally improves access to resources and increases diversification of demand. Finally, the relatively abundant supply of (venture) capital is likely to lower the barriers to develop a new business, and might lure firms (and employees) to capital-rich regions. Even though there is a substantial tradition now on studying the spatial aspects of entrepreneurship, many opportunities for improving insights in this are still not exploited. For example, there are currently many fewer studies on the dynamics in entrepreneurship and complex systems in regional contexts than relatively static studies on the spatial distribution of entrepreneurship in short periods of time. A few research opportunities are discussed, which deal with the nature and measurement of the phenomenon to be explained (entrepreneurship) and the explanatory mechanisms.

Until now, most empirical studies on entrepreneurship have taken new firm formation or self-employment as indicators of entrepreneurship. The advantage of this is that these indicators are widely available in census data or large scale surveys. The disadvantage is that these indicators are too broad, but also too narrow to capture the pursuit of entrepreneurial opportunities. They are *too broad*, because we know that the majority of founders have no motivation to innovate or grow their firm (Santarelli and Vivarelli, 2007; Stam et al., 2008; Stam and Wennberg, 2009; Vivarelli, 2007). This means that future research should use more specific indicators that better capture the phenomenon that is theorized to have an effect on economic development. Traditional (neo-classical) frameworks have been shown to be relevant to analyse the occupational choice between wage earning and self-employment[10] and the entry of new establishments in general. Neo-classical theories have not been able to deal adequately with the innovation and motivations involved in entrepreneurship (see Bianchi and Henrekson, 2005). The latter issues are key in evolutionary economics, and thus provide opportunities for evolutionary economic geography. Different types of entrepreneurship require different types of explanation. Let us take two extremes: the entrepreneur who is self-employed out of necessity, and the ambitious and technologically innovative high-growth start-up. High rates of necessity self-employment are positively affected by high levels of unemployment, which is often driven by decreasing levels of general demand, and maturing industries, and is more likely for low skilled young and old individuals.[11] On the other hand, high rates of ambitious and technologically innovative high-growth entrants are stimulated by high levels of investment in knowledge, an abundant supply of informal investors and venture capital, fluid labour markets and loose buyer–supplier ties. In addition, these firms are most likely to be started by mid-career highly educated (male) individuals. These two explanatory narratives show that entrepreneurship is affected by a multitude of factors, and that different types of entrepreneurship are driven by different mechanisms.

The traditional indicators are also *too narrow*, because there might be entrepreneurship in existing firms ('intrapreneurship') that escapes most data collection efforts in census data and large-scale surveys. This is unfortunate because the effects of these types of entrepreneurship might match or perhaps even overshadow the effects of the majority of independent firms, because intrapreneurs are likely to have much better access to resources, which improves the success chances of their initiatives. They also face a selection environment within their employer's organization that might on the one hand weed out unrealistic market entries, improving the success chance of entrepreneurial initiatives, but on the other hand it might be a too conservative environment, killing off initiatives that might really add value in the long term (but perhaps also cannibalizing the current profit generating activities of the mother firm).[12] In a theoretical perspective, intrapreneurship might be reflected in decisions to set up new product divisions within existing firms, an important element of 'branching processes' as described in Frenken and Boschma (2007). This also connects to a Penrosean theory of firm growth by diversification related to the knowledge base of the firm (Penrose, 1959; Stam et al., 2006).

This discussion of too broad and too narrow indicators of entrepreneurship shows clearly that it does not make sense to talk about *the* geography of entrepreneurship. Different explanatory frameworks are needed to explain different kinds of entrepreneurship. Likewise, existing theoretical frameworks – for example, organizational ecology, social network, and evolutionary economics – do not provide complete explanations for all kinds of entrepreneurship. In order to improve our insights into the spatial variations of entrepreneurship, we need to specify the type of entrepreneurship that is of importance for the research question at hand, so we can match it with the related theoretical framework and empirical indicators.

The studies reviewed share a quantitative, static and rather deterministic orientation, which is in contrast to a more dynamic orientation that is central in evolutionary approaches.[13] In evolutionary approaches the role of agency and interaction with (evolving) selection environments is emphasized. This agency is affected by and affects several ontological layers, ranging from the cognitive abilities of entrepreneurs to macroeconomic and environmental shocks (see Fuller and Moran, 2001). Entrepreneurs are agents who are conditioned by, and sometimes change or even initiate, complex adaptive systems. These systems are situated in particular geographic contexts, and emerge, grow and decline over time. Complex systems that are currently well known because of their high levels of entrepreneurship, for example the high-tech clusters in Silicon Valley and Cambridge, are located in regions that were dominated by agriculture and low levels of entrepreneurship some decades ago. The supply of venture capital is created by co-evolutionary processes in which the emergence of entrepreneurial communities strengthens (for example because of serial entrepreneurs who have sold their businesses and reinvest their money in new ventures as business angels or venture capitalists) and is strengthened by the development of a venture capital community. The latter often follows the emergence of a cluster, not the other way around (Feldman and Francis, 2003; Orsenigo, 2006). In addition, the institutional infrastructure that supports entrepreneurship also often emerges as a product of a critical mass of entrepreneurship in a particular industry or set of related industries. The growth of these industries – by both the indigenous creation of new firms and the attraction of subsidiaries – is then reinforced by this institutional structure (Garnsey and Heffernan, 2005; Keeble et al., 1999). These virtuous cycles of development can turn into vicious circles once congestion effects become stronger and are not offset by the agglomeration economies created in the region. In such a dynamic approach a focus on industry structure is turned into a focus on industrial dynamics. Spatial concentration (or its absence) of an industry is not only an outcome of a process of industrial evolution, but also affects an industry's further evolution (Boschma and Frenken, 2006). This recursive relationship has at least three dimensions (Boschma and Wenting, 2007; Hannan et al., 1995; Stuart and Sorenson, 2003; Van Wissen, 2004). First, geographical concentration of industrial activities can generate agglomeration economies fostering start-ups and innovation and, possibly, the birth of a related industry in the region. Second, geographical concentration of firms increases the level of competition and makes exits of firms raise the average fitness of routines. Third, spatial concentration of firms can also affect the opportunities of collective action as such initiatives are more likely to emerge among proximate agents that can more effectively control opportunistic behaviour.

Most studies on entrepreneurship in evolutionary economics have focused on spin-offs, implicitly focusing on producer entrepreneurship. However, there are indications that user-entrepreneurship is also an important category of entrepreneurship (Shah et al., 2006; Shah and Tripsas, 2007; Von Hippel, 2005). More research on user entrepreneurship is needed, also to acquire better insights into the implications of this phenomenon from the demand side. For example, many early user start-ups in the windsurfing industry were based in Hawaii and California, because these two states, where wind, wave, and temperature conditions made windsurfing a particularly exciting and enjoyable activity, had many early users (Shah, 2005). User entrepreneurship involves entrepreneurship in user communities, but also innovations resulting from differences in individual perceptions (or imaginations) on the part of consumers. The latter relates to individuals' desires for distinction (an important aspect of the cultural industries, and in fashion; Chai et al., 2007) or could emerge out of necessity (to solve particular user problems). Evolutionary economics' focus on the supply side of new variety generation neglects the question of why consumption will not be satiated (see Witt, 2001).

Studies on entrepreneurship in (evolutionary) economic geography are dominated by the firm and the industry as units of analysis. This is in contrast with the focus in entrepreneurship studies: the individual (entrepreneur). Traditionally, evolutionary economic geography regards firms as the agents of change (Boschma, 2004; Frenken and Boschma, 2007), while this chapter has shown that entrepreneurs should at least be incorporated as additional change agents, using firms to pursue the opportunities that they have identified. Future studies should think through and analyse how firms emerge and what role individuals play in this. Individuals have their own (spatial) biographies, which include their prior knowledge and networks that condition the spatial organization of the activities they initiate with their businesses (see Hanson, 2003). Evolutionary economics has often made analogies on the level of the firm (routines, organizational capabilities, Nelson and Winter, 1982; absorptive capacity, Cohen and Levinthal, 1990) that have been derived from individual-level concepts (skills, memory).[14] Take the examples of skills-organizational capabilities. When an individual starts with efforts to exploit an entrepreneurial opportunity, the nature of this opportunity and the probability of successful exploitation is likely to depend on the prior knowledge (Shane, 2000), skills (human capital; Ucbasaran et al., 2006), and networks (Stuart and Sorenson, 2007) of the entrepreneur. Once a successful firm has been created, the importance of the prior knowledge and skills of the entrepreneur is likely to be superseded by the organizational capabilities (including the absorptive capacity) of the firm. These organizational capabilities are more than the sum of the skills or its employees: they are emergent properties. It has been assumed that spin-offs directly inherit routines from their parent (Dahl and Reichstein, 2007; Klepper, 2006), however, this is unlikely as most spin-offs are started by one or a few entrepreneurs, and are thus not started as multi-person organizations that directly inherit or copy the routines from the parent organization. An important empirical question from an entrepreneurship point of view is whether individuals[15] that are able to create a successful firm in one place are also able to create a multi-locational entity that becomes less dependent on their skills, prior knowledge, and to what extent their vision is still driving the development of the firm, for example in overcoming the information distortions that arise in multi-person, multi-locational organizations (see Witt, 1998). Once external shareholders own (parts) of the firm, the question is what kind of consequences this has on the spatial organization of the firm (see Stam, 2006, 2007). The decreased ownership share of the founder (or even exit), makes the firm – and its spatial organization – more sensitive to the capital market as a selection environment.

A relatively unexplored element of entrepreneurship is the exit of (young) firms (see Wennberg, 2009). Firm exit is often regarded as a clear consequence of selection in a product market. However, once we take the aspiration levels and job market opportunities of the entrepreneur into account, the story becomes more complex. Entrepreneurs with relatively low aspiration levels may stay on with their firm for a long time, even though its performance may be regarded as sub-optimal (Gimeno et al., 1997). On the other hand, individuals with relatively high aspiration levels might decide to close down their business even when it is a viable economic entity, because they can achieve a much higher income as a wage earner. These aspiration levels and alternative job markets are likely to be determined by the regional culture and labour market. Until now, we know very little about these mechanisms.

One of the traditional roles of entrepreneurs has been the one of broker between formerly unconnected communities. Bridging two formerly unconnected networks (a so-called structural hole) might provide information benefits to the broker, who might take an entrepreneurial role when he or she exploits the opportunities that are provided with this information (Burt, 2000). Migrant entrepreneurship emphasizes the spatial aspect of this broker role, especially in the recent literature on Argonauts (Saxenian, 2006). More insights into the role of these entrepreneurial migrants in creating and exploiting knowledge flows between particular (place-bound) communities in advanced capitalist economies and developing economies are needed, both on the micro-level processes involved, as well as the effects on the innovativeness and development of the connected (and disconnected) communities. This can be a fertile meeting ground for entrepreneurship studies, international business studies, and economic geography (see Yeung, 2009). More in general, the traditional focus of evolutionary economic geography on advanced capitalist economies has left many opportunities to be pursued for analysing the spatial aspects of entrepreneurship in transition and emerging economies. This road not taken is even more relevant as recent evidence suggests that entrepreneurship is a more important driver of economic growth in transition and emerging economies than in advanced capitalist economies (Stam et al., 2007).

Entrepreneurship is a phenomenon that has become increasingly important as a driver of economic growth in advanced capitalist economies, but also in transition and emerging economies. Economic geography has already built a tradition in analysing the spatial distributions of entrepreneurship (focusing on the regional level of analysis), and evolutionary economics has recently (re)discovered entrepreneurship as a key element of industrial dynamics (focusing on the firm and industry level of analysis), while network approaches have proven to be valuable in the explanation of the behaviour of entrepreneurs. If these approaches are combined in a dynamic perspective and also include the very micro-level of the entrepreneur, the macro-institutional conditions and the levels in between, they will deliver a very powerful approach for understanding wealth creation on multiple levels and multiple time scales.

Acknowledgements

The author wishes to acknowledge comments and suggestions from Ron Boschma, Niels Bosma, Guido Buenstorf, Jan Lambooy, and Ron Martin.

Notes

1. See Breslin (2008) for a comprehensive review of the evolutionary approach to the study of entrepreneurship.
2. Here is a straightforward connection to evolutionary economics, defined as 'that body of economic theory in which the transformation of already existing structures and the emergence and possible spread of novelties are investigated' (Foss, 1994, p. 21).
3. See for the overlaps of evolutionary economics and social network theory Sorenson (2003), of organizational ecology and social network theory Aldrich (1999) and Audia et al. (2006), and of organizational ecology and evolutionary economics Boone and Van Witteloostuijn (1995), Geroski (2001) and Van Wissen (2004).
4. Evolutionary economic geography describes economic development by changes in the time–space distribution of routines (Boschma and Frenken, 2006). This means that the firm rather than the locality is the unit of analysis. The shift from territory to firm resonates a more general reorientation in economic geography from territorial analysis of endowments or institutions to firm analysis of routines and competencies and their embeddedness in the local and global economy (Boschma, 2004; Maskell, 2001; Wrigley et al., 2005).
5. The related concept of competences has even been aggregated from the firm to the regional level (Lawson, 1999).
6. Glaeser (2007) for example found that industry mix and demographic composition (age (older), education (skilled)) are the most important determinants of the heterogeneity in self-employment rates across space.
7. For individuals the probability of *preferring to be self-employed* strongly decreases with age, while the probability of *being self-employed* strongly increases with age (Blanchflower et al., 2001).
8. See also the section on the organizational production of entrepreneurship.
9. A strong science base is not a sufficient condition for an entrepreneurial region to occur. There are multiple regions where a strong scientific base has failed to spawn entrepreneurship (e.g. Ithaca (Cornell) and New Haven (Yale)). A recent study on new firm formation in the Netherlands (Bosma et al., 2007) also did not find a relation between the presence of a university and high levels of new firm formation. The authors explain this as being the result of a general lack of knowledge transfer from universities in the Netherlands.

10. Most clearly demonstrated in situations of rising unemployment: this lowers the opportunity costs for self-employment, other things being equal (Creedy and Johnson, 1983; Evans and Leighton, 1990).
11. As self-employment essentially reflects an occupational choice, the numbers of self-employed might be stimulated by changing labour market regulation and fiscal regimes without any structural change of the economy at all: lowering employment protection and increasing income tax for employees are two institutional changes that are likely to lower the opportunity costs of self-employment. This might have indirect effects on more innovative forms of entrepreneurship as this lowers the inertia of incumbents, but also lowers the incentives of incumbents to invest in their human resources.

    To some degree, self-employment brings us back to a Coasian framework, as the decision for individual actors comes down to whether they join a firm with an employment contract (hierarchy) or transact as a firm on a market.
12. The latter situation is likely to give rise to a spin-off firm.
13. This focus on more static approaches is to some extent explainable by the difficulty of collecting data over longer time spans. There have been more long-term studies in organizational ecology and evolutionary economics, most often based on archival data, but these studies largely focus on single populations/industries, not on whole communities or regions.
14. The traditional focus within evolutionary economics on the firm level concept of routines can be complemented with the individual level concept of skills, and the environmental level concept of institutions (see for example the recent discussion on this in Boschma and Frenken, 2009).
15. This might also involve an entrepreneurial team, adding another unit of analysis in between the individual and the multi-person firm organization.

References

Acs, Z.J. and C. Armington (2004), 'Employment growth and entrepreneurial activity in cities', *Regional Studies*, **38**, 911–27.

Acs, Z.J., D.B. Audretsch, P. Braunerhjelm and B. Carlsson (2005), 'The knowledge spillover theory of entrepreneurship', CEPR Discussion Paper No. 5326, CEPR, London.

Acs, Z.J., N.S. Bosma and R. Sternberg (2008), *The Entrepreneurial Advantage of World Cities*, EIM: Zoetermeer.

Agarwal, R., R. Echambadi, A.M. Franco and M.B Sarkar (2004), 'Knowledge transfer through inheritance: spin-out generation, development and survival', *Academy of Management Journal*, **47,** 501–22.

Aldrich, H. (1999), *Organizations Evolving*, London: Sage.

Anselin, L., A. Varga and Z. Acs (1997), 'Local geographic spillovers between university research and high technology innovations', *Journal of Urban Economics*, **42** (3), 422–48.

Arenius**,** P. and M. Minniti (2005), 'Perceptual variables and nascent entrepreneurship', *Small Business Economics*, **24** (3), 233–47.

Armington, C. and Z.J. Acs (2002), 'The determinants of regional variation in new firm formation', *Regional Studies*, **36** (1), 33–45.

Audia, P.G., J.H. Freeman and P. Reynolds (2006), 'Organizational foundings in community context: instruments manufacturers and their interrelationship with other organizations', *Administrative Science Quarterly*, **51**, 381–419.

Audia, P.G. and C.I. Rider (2005), 'A garage and an idea: what more does an entrepreneur need?', *California Management Review*, **48**, 6–28.

Audia, P.G. and C.I. Rider (2006), 'Entrepreneurs as organizational products revisited', in R. Baum, M. Frese, and R. Baron (eds), *The Psychology of Entrepreneurship*, Magwan, WJ: Lawrence Erlbaum Associates, pp. 113–30.

Audretsch, D.B. and M.P. Feldman (1996), 'R&D spillovers and the geography of innovation and production', *American Economic Review*, **86**, 630–40.

Audretsch, D.B. and M. Fritsch (1994), 'The geography of firm births in Germany', *Regional Studies*, **28** (4), 359–65.

Audretsch, D.B., M.C. Keilbach and E.E. Lehmann (2006), *Entrepreneurship and Economic Growth*, Oxford: Oxford University Press.

Audretsch, D.B. and E. Lehmann (2005a), 'Do university policies make a difference?', *Research Policy*, **34,** 343–47.

Audretsch, D.B. and E. Lehmann (2005b), 'Does the knowledge spillover theory of entrepreneurship hold for regions?', *Research Policy*, **34**, 1191–202.

Audretsch, D.B., E. Lehmann and S. Warning (2005), 'University spillovers and new firm location', *Research Policy*, **34**, 1113–22.

Audretsch, D.B. and P.E. Stephan (1996), 'Company–scientist locational links: the case of biotechnology', *American Economic Review*, **86** (3), 641–52.

Baldwin, J.R., D. Beckstead, W.M. Brown and D.L. Rigby (2008), 'Agglomeration and the geography of localization economies in Canada', *Regional Studies*, **42** (1), 117–32.
Baum, J.A.C. and S.J. Mezias (1992), 'Localized competition and organizational failure: the Manhattan hotel industry 1898–1990', *Administrative Science Quarterly*, **37**, 580–604.
Baum, J.A.C. and B.S. Silverman (2004), 'Picking winners or making them? Alliances, patents, and human capital as selection criteria in venture financing and performance of biotechnology startups', *Journal of Business Venturing*, **19**, 411–36.
Beesley, M.E. and R.T. Hamilton (1994), 'Entry propensity, the supply of entrants and the spatial distribution of business units', *Regional Studies*, **28** (3), 233–39.
Beugelsdijk, S. (2007), 'Entrepreneurial culture, regional innovativeness and economic growth', *Journal of Evolutionary Economics*, **17** (2), 187–210.
Bianchi, M. and M. Henrekson (2005), 'Is neoclassical economics still entrepreneurless?', *Kyklos*, **58** (3), 353–77.
Black, J., D. De Meza and D. Jeffreys (1996), 'House prices, the supply of collateral and the enterprise economy', *Economic Journal*, **106** (434), 60–75.
Blanchflower, D.G., A.J. Oswald and A. Stutzer (2001), 'Latent entrepreneurship across nations', *European Economic Review*, **45** (4–6), 680–91.
Boone, C.A.J.J. and A. Van Witteloostuijn (1995), 'Industrial organization and organizational ecology: the potentials for cross fertilization', *Organization Studies*, **16**, 265–98.
Boschma, R. (2004), 'Competitiveness of regions from an evolutionary perspective', *Regional Studies*, **38** (9), 1001–14.
Boschma, R.A. and K. Frenken (2006), 'Why is economic geography not an evolutionary science? Towards an evolutionary economic geography', *Journal of Economic Geography*, **6** (3), 273–302.
Boschma, R. and K. Frenken (2009), 'Some notes on institutions in evolutionary economic geography', *Economic Geography*, **85** (2), 151–58.
Boschma, R. and R. Wenting (2007), 'Spatial evolution of the British automobile industry: does location matter?', *Industrial and Corporate Change*, **16** (2), 213–38.
Bosma, N. and V. Schutjens (2008), 'Mapping entrepreneurial activity and entrepreneurial attitudes in European regions', *International Journal of Entrepreneurship and Small Business*, **7** (2), 191–213.
Bosma, N.S., A.J. Van Stel and K. Suddle (2007), 'The geography of new firm formation: evidence from independent start-ups and new subsidiaries in the Netherlands', *International Entrepreneurship and Management Journal*, **4** (2), 129–46.
Bottazzi L. and G. Peri (2003), 'Innovation and spillovers in regions: evidence from European patent data', *European Economic Review*, **47**, 687–710.
Braunerhjelm, P. and M. Feldman (2006), *Cluster Genesis. Technology-Based Industrial Development*, Oxford: Oxford University Press.
Brenner, T. and D. Fornahl (2008), 'Regional path dependence in start-up activity', PEEG paper No. T08.12, Urban and Regional Research Centre, Utrecht University.
Breslin, D. (2008), 'A review of the evolutionary approach to the study of entrepreneurship', *International Journal of Management Reviews*, **10** (4), 399–423.
Buenstorf, G. (2007), 'Evolution on the shoulders of giants: entrepreneurship and firm survival in the German laser industry', *Review of Industrial Organization*, **30** (3), 179–202.
Buenstorf, G. and D. Fornahl (2009), 'B2C–bubble to cluster: the dot-com boom, spin-off entrepreneurship, and regional agglomeration', *Journal of Evolutionary Economics*, **19** (3), 349–78.
Buenstorf, G. and S. Klepper (2009), 'Heritage and agglomeration: the Akron tyre cluster revisited', *Economic Journal*, **119** (537), 705–33.
Burt, R.S. (2000), 'The network entrepreneur', in R. Swedberg, (ed.), *Entrepreneurship. The Social Science View*, Oxford: Oxford University Press, pp. 281–307.
Carroll, G. and M.T. Hannan (2000), *The Demography of Corporations and Industries*, Princeton, NJ: Princeton University Press.
Carroll, G. and O. Khessina (2005), 'The ecology of entrepreneurship', in S.A., Alvarez, R. Agarwal and O. Sorenson (eds), *Handbook of Entrepreneurship Research: Disciplinary Perspectives*, New York: Springer, pp. 167–200.
Carroll, G., L. Bigelow, M.D. Seidel and L. Tsai (1996), 'The fates of de novo and de alio producers in the American automobile industry, 1885–1981', *Strategic Management Journal*, **17**, 117–38.
Carter, N.M., W.B. Gartner and P.D. Reynolds (1996), 'Exploring start-up event sequences', *Journal of Business Venturing*, **11** (3), 151–66.
Casper, S. (2007), *Creating Silicon Valley in Europe. Public Policy towards New Technology Industries*, Oxford: Oxford University Press.
Cattani, G., J.M. Pennings and F.C. Wezel (2003), 'Spatial and temporal heterogeneity in founding patterns', *Organization Science*, **14** (6), 670–85.

Chai, A., P.E. Earl and J. Potts (2007), 'Fashion, growth and welfare: an evolutionary approach', *Advances in Austrian Economics*, **10**, 187–207.
Chesbrough, H. (1999), 'The differing organizational impact of technological change: a comparative theory of national institutional factors', *Industrial and Corporate Change*, **8** (3), 447–85.
Chinitz, B. (1961), 'Contrasts in agglomeration: New York and Pittsburg', *American Economic Review*, **51** (2), 279–89.
Cohen, W.M. and D.A. Levinthal (1990), 'Absorptive capacity: a new perspective on learning and innovation', *Administrative Science Quarterly*, **35** (1), 128–52.
Cooper, A.C. and T.B. Folta (2000), 'Entrepreneurship and high technology clusters', in D.L. Sexton and H. Landstrom (eds), *Handbook of Entrepreneurship*, Oxford: Blackwell, pp. 348–67.
Creedy, J. and P.S. Johnson (1983), 'Firm formation in manufacturing industry', *Applied Economics*, **15**, 177–85.
Cross, M. (1981), *New Firm Formation and Regional Development*, Farnborough: Gower.
Dahl, M.S. and T. Reichstein (2007), 'Are you experienced? Prior experience and the survival of new organizations', *Industry and Innovation*, **14** (5), 497–511.
Dahl, M.S. and O. Sorenson (2009), 'The embedded entrepreneur', *European Management Review*, **6**, 172–81.
Davidsson, P. and J. Wiklund (1997), 'Values, beliefs and regional variations in new firm formation rates', *Journal of Economic Psychology*, **18** (2), 179–99.
Dumais, G., G. Ellison and E.L. Glaeser (1997), 'Geographic concentration as a dynamic process', NBER Working Paper No. T 6270.
Evans, D.S. and B. Jovanovic (1989), 'An estimated model of entrepreneurial choice under liquidity constraints', *Journal of Political Economy*, **97**, 808–27.
Evans, L.B. and L.S. Leighton (1990), 'Small business formation by unemployed and employed workers', *Small Business Economics*, **2**, 319–30.
Feldman, M.P. (2001), 'The entrepreneurial event revisited: firm formation in a regional context', *Industrial and Corporate Change*, **10**, 861–91.
Feldman, M.P. and J. Francis (2003), 'Fortune favors the prepared region: the case of entrepreneurship and the capitol region biotechnology cluster', *European Planning Studies*, **11**, 765–88.
Feldman, M.P., J. Francis, and J. Bercovitz (2005), 'Creating a cluster while building a firm: entrepreneurs and the formation of industrial clusters', *Regional Studies*, **39**, 129–41
Figueiredo, O., P. Guimaraes. and D. Woodward (2002), 'Home-field advantage: location decisions of Portuguese entrepreneurs', *Journal of Urban Economics*, **52**, 341–61.
Florida, R. (2002), *The Rise of the Creative Class. And How It's Transforming Work, Leisure, Community, and Everyday Life*, New York: Basic Books.
Forbes, D.P. (1999), 'Cognitive approaches to new venture creation', *International Journal of Management Reviews*, **1** (4), 415–39.
Fornahl, D. (2003), 'Entrepreneurial activities in a regional context', in D. Fornahl and T. Brenner (eds), *Cooperation, Networks and Institutions in Regional Innovation Systems*, Cheltenham, UK and Northampton, MA, USA: Edward Elgar, pp. 38–57.
Foss, N.J. (1994), 'Realism and evolutionary economics', *Journal of Social and Evolutionary Systems*, **17** (1), 21–40.
Fotopoulos, G. and N. Spence (2001), 'Regional variations of firm births, deaths and growth patterns in the UK, 1980–1991', *Growth and Change*, **32**, 151–73.
Freeman, J. (1986), 'Entrepreneurs as organizational products: semiconductor firms and venture capital firms', in G. Libecap (ed.) *Advances in the Study of Entrepreneurship, Innovation, and Economic Growth. A Research Annual. Entrepreneurship and Innovation: the Impact of Venture Capital on the Development of New Enterprise*, Greenwich, CT: JAI Press, pp. 33–52.
Frenken, K. and R.A. Boschma (2007), 'A theoretical framework for evolutionary economic geography: industrial dynamics and urban growth as a branching process', *Journal of Economic Geography*, **7** (5), 635–49.
Frenken K., F. Van Oort and T. Verburg (2007), 'Related variety, unrelated variety and regional economic growth', *Regional Studies*, **41**, 685–97.
Fritsch, M. (2008), 'How does new business formation affect regional development? Introduction to the special issue', *Small Business Economics*, **30** (1), 1–14.
Fritsch, M. and D. Schilder (2008), 'Does venture capital investment really require spatial proximity? An empirical investigation', *Environment and Planning A*, **40** (9), 2114–31.
Fuller, T. and P. Moran (2001), 'Small enterprises as complex adaptive systems: a methodological question?', *Entrepreneurship and Regional Development*, **13** (1), 47–63.
Garnsey E. and P. Heffernan (2005), 'High tech clustering through spin out and attraction; the Cambridge case', *Regional Studies*, **39** (8), 1127–44.
Geroski, P.A. (2001), 'Exploring niche overlaps between organizational ecology and industrial economics', *Industrial and Corporate Change*, **10**, 507–40.

Gibbs, D. (1991), 'Venture capital and regional development: the operation of the venture capital industry in Manchester', *Tijdschrift voor Economische en Sociale Geografie*, **82**, 242–53.
Gimeno, J., T. Folta, A. Cooper and C. Woo (1997), 'Survival of the fittest? Entrepreneurial human capital and the persistence of underperforming firms', *Administrative Science Quarterly*, **42**, 750–83.
Glaeser, E.L. (1999), 'Learning in Cities', *Journal of Urban Economics*, **46**, 254–77.
Glaeser, E.L. (2007), 'Entrepreneurship in the city', NBER Working Paper No. W13551.
Glaeser, E.L. and J.D. Gottlieb (2006), 'Urban Resurgence and the Consumer City', *Urban Studies*, **43** (8), 1275–99.
Glaeser, E.L., J. Kolko and A. Saiz (2000), 'Consumer city', NBER Working Paper 7790.
Gompers, P., J. Lerner and D. Scharfstein (2005), 'Entrepreneurial spawning: public corporations and the genesis of new ventures, 1986 to 1999', *Journal of Finance*, **60** (2), 577–614.
Grebel, T. (2004) *Entrepreneur – A New Perspective*, London: Routledge.
Grebel, T., A. Pyka and H. Hanusch (2003), 'An evolutionary approach to the theory of the entrepreneur', *Industry and Innovation*, **10**, 493–514.
Gudgin, G. (1978), *Industrial Location Processes and Regional Employment Growth*, Westmead: Saxon House.
Hannan, M.T., G.R. Carroll, E.A. Dundon and J.C. Torres (1995), 'Organizational evolution in multinational context: entries of automobile manufacturers in Belgium, Britain, France, Germany and Italy', *American Sociological Review*, **60** (4), 509–28.
Hanson, S. (2003), 'Geographical and feminist perspectives on entrepreneurship', *Geographische Zeitschrift*, **91**, 1–23.
Helfat, C.E. and M.B. Lieberman (2002), 'The birth of capabilities: market entry and the importance of pre-history', *Industrial and Corporate Change*, **11** (4), 725–60.
Hodgson, G.M. and T. Knudsen (2004), 'The firm as an interactor: firms as vehicles for habits and routines', *Journal of Evolutionary Economics*, **14** (3), 281–307.
Holcombe, R. (2007), 'Entrepreneurship and economic growth', in B. Powell (ed.), *Making Poor Nations Rich: Entrepreneurship and the Process of Economic Development*, Stanford, CA: Stanford University Press.
Holtz-Eakin, D., D. Joulfaian and H.S. Rosen (1994), 'Sticking it out: entrepreneurial survival and liquidity constraints', *Journal of Political Economy*, **102** (1), 53–75.
Hoover, E.M. (1948), *The Location of Economic Activity*, New York: McGraw-Hill.
Hoover, E.M. and R. Vernon (1959), *Anatomy of a Metropolis*, Cambridge, MA.: Harvard University Press.
Illeris, S. (1986), 'New firm creation in Denmark: the importance of cultural background', in D. Keeble and E. Wever (eds), *New Firms and Regional Development in Europe*, London: Croom Helm, pp. 141–51.
Jacobs, J. (1969), *The Economy of Cities*, New York: Random House.
Johnson, P.S. and D.G. Cathcart (1979), 'The founders of new manufacturing firms: a note on the size of incubator plants', *Journal of Industrial Economics*, **28** (2), 219–24.
Keeble, D., C. Lawson, B. Moore and F. Wilkinson (1999), 'Collective learning processes, networking and "institutional thickness" in the Cambridge region', *Regional Studies*, **33** (4), 319–32.
Kenney, M. and U. Von Burg (1999), 'Technology and path dependence: the divergence between Silicon Valley and Route 128', *Industrial and Corporate Change*, **8** (1), 67–103.
Kirchhoff, B.A., S.L. Newbert, I. Hasan and C. Armington (2007), 'The influence of university R&D expenditures on new business formations and employment growth', *Entrepreneurship Theory and Practice*, **31** (4), 543–59.
Kirzner, I.M. (1973), *Competition and Entrepreneurship*, Chicago: University of Chicago Press.
Kirzner, I. (1997), 'Entrepreneurial discovery and the competitive market process: an Austrian approach', *Journal of Economic Literature*, **35**, 60–85.
Klepper, S. (2001), 'Employee startups in high-tech industries', *Industrial and Corporate Change*, **10** (3), 639–74.
Klepper, S. (2002), 'The capabilities of new firms and the evolution of the US automobile industry', *Industrial and Corporate Change*, **11** (4), 645–66.
Klepper, S. (2006), 'The evolution of geographic structure in new industries', *Revue de l'OFCE* 2006- 5 (97.2), 135–58.
Klepper, S. (2007), 'Disagreements, spinoffs, and the evolution of Detroit as the capital of the U.S. automobile industry', *Management Science*, **53**, 616–31.
Klepper, S. and K. Simons (2000), 'Dominance by birthright: entry of prior radio producers and competitive ramifications in the U.S. television receiver industry', *Strategic Management Journal*, **21** (10–11), 997–1016.
Klepper, S. and S. Sleeper (2005), 'Entry by spin-off', *Management Science*, **51** (8), 1291–306.
Kristensen, P.H. (1994), 'Spectator communities and entrepreneurial districts', *Entrepreneurship and Regional Development*, **6**, 177–98.
Lafuente, A. and V. Salas (1989), 'Types of entrepreneurs and firms: the case of new Spanish firms', *Strategic Management Journal*, **10**, 17–30.

Lawson, C. (1999), 'Towards a competence theory of the region', *Cambridge Journal of Economics*, **23** (2), 151–66.
Lee, S.Y., R. Florida and Z.J. Acs (2004), 'Creativity and entrepreneurship: a regional analysis of new firm formation', *Regional Studies*, **38** (8), 879–91.
Lloyd, P. and C. Mason (1984), 'Spatial variation in new firm formation in the United Kingdom: comparative evidence from Merseyside, Greater Manchester and South Hampshire', *Regional Studies*, **18**, 207–20
McFadden, D. (2001), 'Economic choices', *American Economic Review*, **91**, 351–78.
Malmberg, A. and P. Maskell (2002), 'The elusive concept of localisation economies – towards a knowledge-based theory of spatial clustering', *Environment and Planning A*, **34**, 429–49.
Marlet, G. and C. Van Woerkens (2007), 'The Dutch creative class and how it fosters urban employment growth', *Urban Studies*, **44** (13), 2605–26.
Marshall, A. (1890), *Principles of Economics*, London: Macmillan.
Martin, R. and P. Sunley (2003), 'Deconstructing clusters: chaotic concept or policy panacea?', *Journal of Economic Geography*, **3**, 5–35.
Martin, R., C. Berndt, B. Klagge and P. Sunley (2005), 'Spatial proximity effects and regional equity gaps in the venture capital market: evidence from Germany and the United Kingdom', *Environment and Planning A*, **37**, 1207–31.
Martin, R., P. Sunley and D. Turner (2002), 'Taking risks in regions: the geographical anatomy of Europe's emerging venture capital market', *Journal of Economic Geography*, **2**, 121–50.
Maskell, P. (2001), 'The firm in economic geography', *Economic Geography*, **77** (4), 329–44.
Mason, C.M. (1991), 'Spatial variations in enterprise', in R. Burrows (ed.), *Deciphering the Enterprise Culture*, London: Routledge, pp. 74–106.
Mason, C.M. and R.T. Harrison (1999), 'Financing entrepreneurship: venture capital and regional development', in R. Martin (ed.), *Money and the Space Economy*, Chichester: Wiley, pp. 157–83.
Mason, C.M. and R.T. Harrison (2002), 'The geography of venture capital investments in the UK', *Transactions of the Institute of British Geographers*, **27**, 427–51.
Metcalfe, J.S. (2002), 'Growth of knowledge and knowledge of growth', *Journal of Evolutionary Economics*, **12**, 3–15.
Metcalfe, J.S. (2004), 'The entrepreneur and the style of modern economics', *Journal of Evolutionary Economics*, **14** (2), 157–75.
Michelacci, C. and O. Silva (2007), 'Why so many local entrepreneurs?', *Review of Economics and Statistics*, **89** (4), 615–33.
Minniti, M. (2005), 'Entrepreneurship and network externalities', *Journal of Economic Behavior and Organization*, **57** (1), 1–27.
Moore, G. and K. Davis (2004), 'Learning the Silicon Valley way', in T. Bresnahan and A. Gambardella (eds), *Building High-Tech Clusters: Silicon Valley and Beyond*, Cambridge: Cambridge University Press, pp. 7–39.
Mueller, P. (2006), 'Entrepreneurship in the region: breeding ground for nascent entrepreneurs?', *Small Business Economics*, **27**, 41–58
Nanda, R. and J.B. Sørensen (2008), 'Peer effects and entrepreneurship', HBS Working Paper 08–051.
Nelson, R.R. and S.G. Winter (1982), *An Evolutionary Theory of Economic Change*, Cambridge, MA: The Belknap Press of Harvard University Press.
Niittykangas, H. and H. Tervo (2005), 'Spatial variations in intergenerational transmission of self-employment', *Regional Studies*, **39** (3), 319–32.
Nooteboom, B. (2000), *Learning and Innovation in Organizations and Economies*, Oxford: Oxford University Press.
Nooteboom, B. (2008), 'Learning, discovery and collaboration', in B. Nooteboom, and E. Stam (eds), *Micro-Foundations for Innovation Policy*, Amsterdam/Chicago: Amsterdam University Press/Chicago University Press, pp.75–102.
O'Farrell, P.N. and R. Crouchley (1984), 'An industrial and spatial analysis of new firm formation in Ireland', *Regional Studies*, **18**, 221–36.
Orsenigo, L. (2006), 'Clusters and clustering: stylized facts, issues, and theories', in: P. Braunerhjelm, and M. Feldman (eds), *Cluster Genesis. Technology-Based Industrial Development*, Oxford: Oxford University Press, pp. 195–218.
Parwada, J.T. (2008), 'The genesis of home bias? The location and portfolio choices of investment company start-ups', *Journal of Financial and Quantitative Analysis*, **43**, 245–66.
Penrose, E.T. (1959), *The Theory of the Growth of the Firm*, 3rd edn (1995), Oxford: Oxford University Press.
Phillips, D.J. (2002), 'A genealogical approach to organizational life chances: the parent–progeny transfer among Silicon Valley law firms, 1946–1996', *Administrative Science Quarterly*, **47** (3), 474–506.
Piore, M.J. and C.F. Sabel (1984), *The Second Industrial Divide: Possibilities for Prosperity*, New York: Basic Books.

Powell, W.W., K. Koput, J.L. Bowie and L. Smith-Doerr (2002), 'The spatial clustering of science and capital: accounting for biotech firm-venture capital relationships', *Regional Studies*, **36** (3), 291–306.
Pred, A. (1967), *Behavior and Location: Foundations for a Geographic and Dynamic Location Theory*, Part I, Lund Studies in Geography B 27, Lund, Sweden. University of Lund.
Reynolds, P., D.J. Storey and P. Westhead (1994), 'Cross-national comparisons of the variation in new firm formation rates', *Regional Studies*, **28**, 443–56.
Reynolds, P.D. and S.B. White (1997), *The Entrepreneurial Process: Economic Growth, Men, Women, and Minorities*, Westport, CT: Quorum Books.
Rocha, H.O. (2004), 'Entrepreneurship and development: the role of clusters. a literature review', *Small Business Economics*, **23**, 363–400.
Santarelli, E. and M. Vivarelli (2007), 'Entrepreneurship and the process of firms' entry, survival and growth', *Industrial and Corporate Change*, **16**, 455–88.
Sarasvathy, S., N. Dew, S. Velamuri and S. Venkataraman (2003), 'Three views of entrepreneurial opportunity', in Z.J. Acs and D.B. Audretsch (eds), *Handbook of Entrepreneurship Research. An Interdisciplinary Survey and Introduction*, Boston, MA: Kluwer Academic Publishers, pp.141–60.
Saxenian, A. (1990), 'Regional networks and the resurgence of Silicon Valley', *California Management Review*, **33** (1), 89–112.
Saxenian, A. (1994), *Regional Advantage: Culture and Competition in Silicon Valley and Route 128*, Cambridge, MA: Harvard University Press.
Saxenian, A. (2006), *The New Argonauts. Regional Advantage in a Global Economy*, Cambridge, MA: Harvard University Press.
Savage, M., P. Dickens and T. Fielding (1988), 'Some social and political implications of the contemporary fragmentation of the service class in Britain', *International Journal of Urban and Regional Research*, **12** (3), 455–76.
Schumpeter, J.A. (1934), *The Theory of Economic Development*, Cambridge, MA: Harvard University Press.
Schumpeter, J.A. (1942), *Capitalism, Socialism, and Democracy*, New York: Harper & Brothers.
Shah, S. (2005), 'From innovation to firm formation in the windsurfing, skateboarding, and snowboarding industries', *University of Illinois Working Paper* No. T 05–0107.
Shah, S. and M. Tripsas (2007), 'The accidental entrepreneur: the emergent and collective process of user entrepreneurship', *Strategic Entrepreneurship Journal*, **1** (1), 123–40.
Shah, S., R. Agarwal and D. Audretsch (2006), 'The knowledge context and the entrepreneurial process: academic, user, and employee entrepreneurship', University of Illinois Working Paper.
Shane, S. (2000), 'Prior knowledge and the discovery of entrepreneurial opportunities', *Organization Science*, **11**, 448–69.
Simon, H. (1957), *Administrative Behavior*, 2nd edn, New York: Macmillan.
Sørensen, J. (2007), 'Bureaucracy and entrepreneurship: workplace effects on entrepreneurial entry', *Administrative Science Quarterly*, **52**, 387–412.
Sorenson, O. (2003), 'Social networks and industrial geography', *Journal of Evolutionary Economics*, **13**, 513–27.
Sorenson, O. (2005), 'Social networks and the persistence of clusters: evidence from the computer workstation industry', in S. Breschi and F. Malerba (eds), *Clusters, Networks, and Innovation*, Oxford: Oxford University Press, pp. 297–316.
Sorenson, O. and P.G. Audia (2000), 'The social structure of entrepreneurial activity: geographic production of footwear in the United States, 1940–1989', *American Journal of Sociology*, **106**, 1546–88.
Sorenson, O. and T. Stuart (2001), 'Syndication networks and the spatial distribution of venture capital investments', *American Journal of Sociology*, **106**, 1546–88.
Stam, E. (2006), 'A process model of locational change: implications for an evolutionary economic geography', in A. Pyka & H. Hanusch (eds), *Applied Evolutionary Economics and the Knowledge-Based Economy*, Cheltenham, UK and Northampton, MA, USA: Edward Elgar, pp. 143–65.
Stam, E. (2007), 'Why butterflies don't leave. Locational behavior of entrepreneurial firms', *Economic Geography*, **83** (1), 27–50.
Stam, E. (2008), 'Entrepreneurship and innovation policy', in B. Nooteboom and E. Stam (eds), *Micro-Foundations for Innovation Policy*, Amsterdam/Chicago: Amsterdam University Press/Chicago University Press, pp. 135–72.
Stam, E. and K. Wennberg (2009), 'The roles of R&D in new firm growth', *Small Business Economics*, **33** (1), 77–89.
Stam, E., E. Garnsey and P. Heffernan (2006), 'A Penrosean theory of the firm: implications and applications for new firm growth', in M. Dietrich (ed.), *Economics of the Firm: Analysis, Evolution and History*, London: Routledge, pp. 146–68.
Stam, E., P. Gibcus, J. Telussa and E. Garnsey (2008), 'Employment growth in new firms' first decade', in H. Landström (ed.), *Entrepreneurship, Sustainable Growth and Performance: Frontiers in*

European Entrepreneurship Research, Cheltenham UK and Northampton, MA, USA: Edward Elgar, pp. 324–51.
Stam, E., K. Suddle, S.J.A. Hessels and A. Van Stel (2007), 'High growth entrepreneurs, public policies and economic growth', Jena Economic Research Papers 2007–019, Friedrich-Schiller-University Jena/Max-Planck-Institute of Economics.
Storey, D.J. (1982), *Entrepreneurship and the New Firm*, London: Croom Helm.
Stuart, T.E. and O. Sorenson (2003), 'The geography of opportunity: spatial heterogeneity in founding rates and the performance of biotechnology firms', *Research Policy*, **32**, 229–53.
Stuart, T.E. and O. Sorenson (2007), 'Strategic networks and entrepreneurial ventures', *Strategic Entrepreneurship Journal*, **1**, 211–27.
Ucbasaran, D., P. Westhead and M. Wright (2006), *Habitual Entrepreneurs*, Cheltenham, UK and Northampton, MA, USA: Edward Elgar.
Uhlaner, L.M. and A.R. Thurik (2007), 'Postmaterialism influencing total entrepreneurial activity across nations', *Journal of Evolutionary Economics*, **17** (2), 161–85.
Vaillant, Y. and E. Lafuente (2007), 'Do different institutional frameworks condition the influence of local fear of failure and entrepreneurial examples over entrepreneurial activity?', *Entrepreneurship and Regional Development*, **19**, 313–37.
Van Aalst, I., O. Atzema, R. Boschma and F.G. Van Oort (2006), 'Creatieve klasse en regionaal-economische groei', [The creative class and economic growth], in S. Raes and B. Hofstede (eds), *Creatief vermogen. De economische potentie van cultuur en creativiteit* [Creative Power: The Economic Potential of Culture and Creativity], The Hague: Elsevier, pp. 123–54.
Van Wissen, L. (2004), 'A spatial interpretation of the density dependence model in industrial demography', *Small Business Economics*, **22** (3–4), 253–64.
Varga, A. (1998), 'Local academic knowledge spillovers and the concentration of economic activity', Research Paper No. 9803, Regional Research Institute, West Virginia University, Morgantown.
Vivarelli, M. (1991), 'The birth of new enterprises', *Small Business Economics*, **3**, 215–23.
Vivarelli, M. (2007), *Entry and Post-Entry Performance of Newborn Firms*, London: Routledge.
Von Hippel, E. (2005), *Democratizing Innovation*, Cambridge, MA: MIT Press.
Wagner, J. and R. Sternberg (2004), 'Start-up activities, individual characteristics, and the regional milieu: lessons for entrepreneurship support policies from German micro data', *Annals of Regional Science*, **38**, 219–40.
Wennberg, K. (2009), *Entrepreneurial Exit*, Stockholm: Stockholm School of Economics.
Wennberg, K.J., T. Folta and F. Delmar (2007), 'A real options model of stepwise entry into self-employment', The Office of Advocacy Small Business Working Papers 07, Washington.
Wenting, R. (2008), 'Spinoff dynamics and the spatial formation of the fashion design industry, 1858–2005', *Journal of Economic Geography*, **8** (5), 593–614.
Wezel, F.C. (2005), 'Location-dependence and industry evolution. Founding rates in the United Kingdom motorcycle industry, 1895–1993', *Organization Studies*, **26** (5), 729–54.
Wezel, F.C., G. Cattani and J.M. Pennings (2006), 'Competitive implications of interfirm mobility', *Organization Science*, **17** (6), 691–709.
Witt, U. (1998), 'Imagination and leadership – the neglected dimension of an evolutionary theory of the firm', *Journal of Economic Behavior and Organization*, **35** (2), 161–77.
Witt, U. (2001), 'Learning to consume: a theory of wants and the growth of demand', *Journal of Evolutionary Economics*, **11**, 23–36.
Witt, U. (2003), *The Evolving Economy. Essays on the Evolutionary Approach to Economics*, Cheltenham, UK and Northampton, MA, USA: Edward Elgar.
Witt, U. (2005), 'On novelty and heterogeneity', in T. Lux, S. Reitz and E. Samanidou (eds), *Nonlinear Dynamics and Heterogeneous Interacting Agents*, Heidelberg: Springer, pp. 123–38.
Woodward, D., O. Figueiredo and P. Guimaraes (2006), 'Beyond the Silicon Valley: university R&D and high-technology location', *Journal of Urban Economics*, **60** (1), 15–32.
Wrigley, N., N.M. Coe and A.D. Currah (2005), 'Globalizing retail: conceptualizing the distribution-based TNC', *Progress in Human Geography*, **29** (4), 437–57.
Yeung, H.W. (2009), 'Transnationalizing entrepreneurship: a critical agenda for economic geography', *Progress in Human Geography*, **33** (2), 210–35.
Zander, I. (2004), 'The microfoundations of cluster stickiness – walking in the shoes of the entrepreneur', *Journal of International Management*, **10**, 151–75.
Zook, M.A. (2002), 'Grounded capital: venture financing and the geography of the internet industry, 1994–2000', *Journal of Economic Geography*, **2** (2), 151–77.
Zucker, L.G., M.R. Darby and M.B. Brewer (1998), 'Intellectual human capital and the birth of U.S. biotechnology enterprises', *American Economic Review* **88**, 290–306.

7 Pecuniary externalities and the localized generation of technological knowledge[1]

Cristiano Antonelli

1. Introduction

The literature on knowledge spillovers has been growing exponentially in the recent years. This literature has elaborated the Marshallian understanding of technological externalities where knowledge is a production factor spilling into the atmosphere of industrial districts, at no costs for perspective users either to acquire or to use it: knowledge can be acquired with no transaction and communication costs.

The new growth theory has further elaborated this literature with the distinction between generic and specific technological knowledge. Generic technological knowledge is germane to a variety of uses, while specific technological knowledge is embodied in production processes and routines: as such it has strong idiosyncratic features. 'Inventors' can appropriate specific knowledge; generic knowledge instead retains the typical features of the Arrovian public good. The appropriability of specific knowledge provides sufficient incentives for investment in knowledge-generating activities. The assumption about the intrinsic complementarity between generic and specific knowledge is the basic engine of the process. Innovators generate generic knowledge while they are engaged in the introduction of new specific knowledge embodied in new products and new processes. The production of specific knowledge takes advantage of the collective availability of generic knowledge. The spillover of generic knowledge helps the generation of new specific knowledge by third parties and yet does not reduce the incentives to the generation of new knowledge for the strong appropriability of the specific applications. Each firm has unlimited access to the spillovers of generic knowledge that can be used with no efforts. According to the new growth theory, the unconditional and unconstrained access to generic technological knowledge leads to the spontaneous and ubiquitous increase of total factor productivity and hence the automatic growth of output (Jones, 2002; Romer, 1990).

Recent acquisitions of the applied economics of knowledge have shown that the access to external knowledge requires dedicated resources and entails specific activities: knowledge does not spill freely into the atmosphere (Cohen and Levinthal, 1989 and 1990; Griffith et al., 2003). In this context it seems appropriate to reconsider the notion of knowledge externalities.

As a matter of fact the Marshallian analysis of the notion of externalities is much more widely discussed. Two quite different types of externalities have been identified in the Marshallian literature: (1) technological externalities, and (2) pecuniary externalities. Technological externalities take place when unpaid production factors enter the production function of users via direct interactions that are not mediated by the price mechanisms. Pecuniary externalities apply when the prices of both products and factors differ from equilibrium levels, because of the effects of external forces. So far they have found little application in the economics of knowledge, as the literature has explored more systematically the consequences of knowledge non-appropriability in terms of 'direct interdependence' non-mediated by the market mechanism. It seems in fact clear that the new growth theory and the large empirical literature initiated by Zvi Griliches (1979 and 1992), with the notion of technological spillovers, elaborate on the notion of 'technological externalities'.

This chapter explores an alternative analytical path, based on the notion of pecuniary knowledge externalities. Because pecuniary knowledge externalities apply to studies not only of production functions but also of cost and revenue functions, they provide a novel and fruitful tool that makes it possible to identify both negative and positive effects of the external context, their interplay and hence the dynamics of the generation of technological knowledge. Hence they help the construction of an evolutionary economic geography (Boschma and Frenken, 2006; Boschma and Martin, 2007).

2. The role of external factors in the localized generation of technological knowledge

In the Arrovian approach, technological knowledge is viewed as the result of a top-down process (Arrow, 1962a). Scientific knowledge is generated in universities in the form of general principles and universal laws. Eventually it may be applied to production processes so as to feed the generation of new technological knowledge. The bottom-up approach according to which learning is the main, if not the exclusive, source of knowledge has challenged the Arrovian approach. The resource-based theory of the firm has provided the foundations for this approach and has highlighted the key role played by learning routines in the generation of knowledge. In this approach to the economics of knowledge the distinction between tacit and codified knowledge plays a key role.

The resource-based theory of the firm presents learning as the joint product of current activities and hence assimilates knowledge to learning. Edith Penrose (1959 [1995]) identifies the firm, its organization and its routines, as the privileged actor in the learning process. The firm precedes the production function as its primary activity consists of the generation of new technological knowledge. Each firm, as is well known, learns and builds up new capabilities and eventually discovers new possible applications for production factors and competences that are found within its own boundaries. According to the resource-based theory, in other words, innovative firms are successful when they try to make the most effective use of production factors that are not only locally abundant, but also internally – within its own boundaries – abundant (Foss, 1997 and 1998). The bottom-up approach to understanding the dynamics of knowledge highlights the role of internal learning processes, as the necessary and sufficient condition for the generation of new knowledge at large (Foss and Mahnke, 2000).

The economics of localized technological change make it possible to implement the resource-based theory of the firm and hence bring a bottom-up approach to the economics of knowledge in two cardinal ways: (1) qualifying the conditions for generation as being shaped by the localization of the learning process; (2) emphasizing the contribution made by intentional decision making that stems from the creative reaction of innovative firm. Let us analyse them in turn.

The Penrosean analysis of the key role of learning in the generation of technological knowledge and in the introduction of technological changes has been greatly qualified and sharpened by the insight of Anthony Atkinson and Joseph Stiglitz (1969) who introduced the strong hypothesis that technological change can take place only in a limited technical space, defined in terms of factor intensity (Arrow, 1962b; Penrose, 1959). Technological change is localized because it has limited externalities and affects only a limited range of the techniques, that is those contained by a given isoquant, which is identified by the actual context of learning, in the proximity of equilibrium conditions where the firms have been producing. In other words technological change can only take place where firms have been able to learn: localization here is strictly defined in terms of factor intensity and with respect to the techniques in place at each point in time.

The localized approach paves the way to implementing a broader understanding of what determines and conditions the generation of technological knowledge. The notion of localized technological knowledge in fact makes it possible to emphasize the role of knowledge as a joint-product of the economic and productive activity. Agents learn how, when, where and what, also and mainly through experience, accumulated in daily routines. Firms, however, can also generate new technological knowledge by means of research and development activities: learning is not the single input into the generation of new knowledge (Antonelli, 1995 and 1999).

The introduction of new technologies is constrained by how much competence and experience has been accumulated through learning processes in specific technical and contextual procedures. Agents, in this approach, can generate new knowledge only in limited domains and fields where they have accumulated sufficient levels of competence and experience. A strong complementarity must be assumed between learning, as a knowledge input, and other knowledge inputs such as R&D laboratories, within each firm (Antonelli, 2001).

Learning indeed is one of the basic sources of new technological knowledge. As such it exerts a strong and clear effect in terms of defining the cognitive space into which each firm can expand its current technological base. As a consequence the new technological knowledge generated by each firm is constrained within the proximity of its current activities. In other words, learning exerts a powerful localizing effect, which limits the spectrum of possible discoveries. At the same time however the generation of new knowledge can take a wide variety of possible directions impinging on both the specific form of learning that is actively implemented and the context in which it takes place (Antonelli, 1995, 1999, 2001).

The transformation of a competence based on learning processes into new, actual technological knowledge requires specific and dedicated effort. The generation of new technological knowledge can be considered to be the specific activity of the firm and its distinctive function within the economic system: the firm is indeed the locus of technological discovery. Yet discovery and creativity are not automatic, incremental, past dependent and hence deterministic activities guided by the sheer accumulation of internal competence based on tacit learning, but rather the result of a complex path-dependent process where at each point in time firms make explicit and intentional efforts to generate new technological knowledge.[2] Such efforts are most likely to be successful when a number of contextual and external conditions apply (Antonelli, 2007).

3. External knowledge as a production factor

In order to generate new knowledge, firms need to combine and integrate internal sources of knowledge such as intramural research and development activities and learning processes, with the systematic use of external knowledge as a primary input for the general production of new knowledge. No firm, in fact, can innovate in isolation. External knowledge is an essential input into the generation of new knowledge. External knowledge can substitute internal sources of knowledge only to a limited extent: full-fledged substitutability between internal and external knowledge cannot apply. Unconstrained complementarity however also appears inappropriate. Building on the large empirical evidence about the role of external knowledge, the hypothesis of a constrained multiplicative relationship can be articulated. External and internal knowledge, both in their tacit and codified form, are complementary inputs where none is disposable. The ratio of internal to external knowledge however seems relevant. Neither can firms generate new knowledge relying only on external or internal knowledge as the single input. With an appropriate ratio of internal to external knowledge instead, internal knowledge and external knowledge inputs enter into a constrained multiplicative production function. Both below and above the threshold of the appropriate combination of the complementary inputs the firm cannot achieve the maximum output (Patrucco, 2009).

Because of the intrinsic indivisibility of technological knowledge, the successful generation of new knowledge depends on the access to external knowledge. External knowledge is only potentially useful: systematic efforts have to be made in order to take advantage of such possibilities. To do so, firms rely on knowledge exploration strategies to identify the sources of knowledge, to assess whether and how to rely on external or internal knowledge in the production of new knowledge. Only when a firm is able to fully coordinate learning and research activities conducted within its boundaries with the relevant sources of external knowledge, both tacit and codified, can new knowledge be successfully generated. Knowledge procurement is as relevant as intramural research activities in the generation of new knowledge. The purchase of patents and licenses in knowledge markets by means of knowledge transactions, however, is by no means the single source of external knowledge. External knowledge can be accessed also by means of a variety of other tools, including the hiring of qualified personnel embodying the competence acquired by means of learning in other companies and an array of interaction modes with public research centers, customers, suppliers and competitors.

The acquisition of external knowledge is expensive both in terms of actual purchasing costs and in terms of knowledge governance costs. Knowledge governance costs include all knowledge transaction, communication and networking costs. Knowledge transaction costs are the costs associated with the exploration activities in the markets for disembodied knowledge such as search, screening, processing, and contracting. As it is well known the assessment of the actual quality of knowledge can be difficult when the vendor bears the risks of opportunistic behavior and dangerous disclosure.

The acquisition of external knowledge requires qualified interactions with other agents: dedicated efforts are necessary to create the institutional context into which external knowledge can be acquired. The capability of agents to access external technological knowledge depends on the fabric of institutional relations and shared codes of understanding that help to reduce information asymmetries, reducing the scope for opportunistic behavior and building a context into which reciprocity, constructed trust and generative relationship can be implemented (Cohen and Levinthal, 1989 and 1990; Griffith et al., 2003).

Knowledge communication is necessary when knowledge is dispersed and fragmented, retained by a myriad of heterogeneous agents, and yet characterized by high levels of indivisibility with important potential benefits in terms of externalities stemming from its integration and recombination. Yet knowledge communication is not automatic. On the contrary, it is the result of much intentional activity designed to create a context conducive to combining variety and complementarity.

Systematic networking is necessary to establish knowledge communication flows. The network structure of the system plays a key role in shaping the flows of knowledge communication and hence the availability of external knowledge. Specific, dedicated networking activities are necessary in order to manage the flows of knowledge that are not internal to each firm and yet cannot be reduced to arm's length transactions. Networking activities make knowledge interactions, as distinct from knowledge transactions, possible. Networking activities are a specific – indispensable – ingredient of the basic governance of knowledge (Freeman, 1991).

Firms often rely on networking interactions with other independent parties, to increase the proprietary control of their knowledge, to acquire external knowledge and to better exploit it. External knowledge can be acquired by taking advantage of the spillovers from the academic activities, and from localization in the proximity of other firms. Qualified user–producer interactions, both upstream, with suppliers, and downstream, with customers, are the source of key inputs into the production of new knowledge. Knowledge search and utilization is better implemented within networks of interactions based on constructed and repeated interactions, qualified by contractual relations. The array of networking tools is ever increasing and includes both formal and informal mechanisms. Joint ventures, dedicated research clubs, sponsored spin-offs, patent-thicketing, technological platforms, cross-licensing, and in-house outsourcing are the main types of formal cooperative tool. Co-localization within technological districts and membership of epistemic communities are typical forms of networking procedures (Antonelli, 2008a).

Our basic assumption here is that the levels of knowledge governance costs have a key role in assessing the actual levels of the total costs for the prospective users of external knowledge (Arrow, 1969).

The understanding of the costs of external knowledge has important implications regarding the direction and the amount of technological knowledge being generated by the firm. When efficient markets for knowledge are available, the selection of knowledge activities that firms retain within their boundaries is much more effective. The search for the inclusion of knowledge-generating activities and their eventual valorization is in fact much more selective. The exploration for external sources of knowledge and knowledge outsourcing becomes common practice. Firms can rely on external providers for specific bits of complementary knowledge. Knowledge outsourcing on the demand side matches the supply of specialized knowledge-intensive business service firms. Universities and other public research centers can complement their top-down research activities finalized to the production of scientific knowledge with the provision of elements of technological knowledge to business firms.

The stronger are pecuniary knowledge externalities, the stronger the incentives for firms to select the characteristics of the technological knowledge they can generate, according to the characteristics of the context into which they are embedded. A variety of factors affect this process: the cognitive distance among agents, the complementarity in competence and research agendas, the levels of trust, and the institutional setting. Geographic proximity plays a key role (Boschma, 2005).

Firms that have access to cheaper external knowledge, can generate a larger amount of knowledge with a given amount of resources available to fund research activities. The unit costs of knowledge generated in a conducive environment are clearly lower than the unit costs of the knowledge generated in a 'hostile' context by a single firm able to rely almost exclusively on its own internal competence.

This analysis has many important implications for the role of the local context into which firms are embedded. It is clear, for instance, that when and where external knowledge is cheap, both because of low purchasing costs in the markets for codified knowledge, and low knowledge governance costs, firms will rely less on internal learning and research activities. On the contrary, when and where the access conditions to external knowledge are less easy, firms will rely more on internal research and learning activities. This analysis provides a clue to understanding the puzzling evidence about the low levels of formal research activities of firms localized in fertile and dynamic technological districts (Antonelli, 2008b).

4. Pecuniary knowledge externalities

The new evidence provided by the applied economics of knowledge about the costs of acquisition of external knowledge (Cohen and Levinthal, 1989 and 1990), the identification of the dual role of technological knowledge, as both an input and output, elaborated by David (1993), and the new understanding about technological knowledge as a distributed factor (Hayek, 1945; Martin, 1999 and 2007) have brought important changes to the theory. Let us consider them in turn.

The acquisition of technological knowledge requires some dedicated resources. Technological knowledge spills into the atmosphere, but its use entails some costs. Imitation costs are relevant as much as knowledge governance costs in terms of transaction, interaction and communication costs. Because of the intrinsic non-exhaustibility of knowledge, however, the costs of existing knowledge are far below the costs of its generation. Even after the proper assessment of knowledge governance costs it becomes more and more evident that their levels can be lower than the costs of early generation, at least in specific and positive geographic, historic, institutional and sectoral contexts (Antonelli, 2007 and 2008a).

The understanding of the key role of knowledge as an input in the production of new knowledge (David, 1993) adds new elements to understanding the intrinsic complementarity between external and internal sources of knowledge for the production of new knowledge. The legacy of Hayek (1945) finds new support: technological knowledge is viewed as dispersed and fragmented into a variety of complementary and yet specific and idiosyncratic applications and contexts.

In such a new context, where knowledge is viewed as a collective activity, the application of the notion of pecuniary externalities to the economics of knowledge makes it possible:

1. To qualify the systemic characteristics that favor the generation of technological knowledge. Agglomeration within technological systems, both in geographical and technological space, favors the generation of new knowledge only in specific contexts

where and when positive knowledge externalities that knowledge as an input makes available at costs that are lower than equilibrium levels are not offset by negative externalities that reduce the price that knowledge as an output can command in market exchanges. Such circumstances in fact do not hold everywhere and at all times, but only in highly idiosyncratic conditions (Antonelli, 2008b); and
2. To appreciate the negative effects of excess proximity within geographical and technological clusters in terms of both reduced levels of knowledge appropriability and hence reduction of the prices for the products that embody new proprietary knowledge, and increased knowledge governance costs.

These elements make it possible to reconsider the notion of externalities. As it is well known, the Marshallian literature has identified two quite different types of externalities: (1) technological externalities, and (2) pecuniary externalities. Technological externalities consist of direct interdependence among producers. Pecuniary externalities consist of indirect interdependence. In the former case the interdependence is not mediated by the market mechanisms. In the latter, interdependence takes place via effects on the price system (Martin, 2007; Meade, 1952; Scitovsky, 1954; Viner, 1931).

Technological externalities exist when unpaid production factors enter the production function. According to Scitovsky this is the case of technological external economies. They apply when:

> The producer's output may be influenced by the action of persons more directly and in other ways than through their offer of services used and demand for products produced by the firm. This is the counterpart of the previous case, and its main instance is inventions that facilitate production and become available to producers without charge. (Scitovsky, 1954, p. 144)

Pecuniary externalities apply when the prices for production factors differ from equilibrium levels and reflect the effects of external forces. According to Scitovsky (1954), pecuniary external economies consist of 'interdependence among producers through the market mechanism' (p. 146).[3] There are positive pecuniary externalities when the market price of production factors happens to be lower than equilibrium levels because of the effects of market interactions among firms in the growth process.

So far, the notion of pecuniary externalities has found little application in understanding the *generation* of technological change, as distinct from the adoption and direction of technological knowledge, and as distinct from the rate, of technological change. Pecuniary knowledge externalities become relevant as soon as: (1) firms are credited with the creative capability to generate intentionally technological knowledge and to introduce technological changes that are consistent with their specific and contextual conditions, and (2) the active role of knowledge users is recognized. In order to command new technological knowledge generated by third parties, and to take advantage of it, users need to perform specific activities that entail specific resources. This is true for the adopters and imitators of new products and processes when technological knowledge is embodied, for the customers of patents and licenses when knowledge is disembodied, and for the preceptors of knowledge spillovers. In all cases users can access external knowledge only at a cost: such costs have an effect on a firm's technological choices. Hence the need for pecuniary knowledge externalities (David and Rosenbloom, 1990).

The appreciation of pecuniary externalities makes it possible to understand how the lower relative prices of both specific knowledge inputs and knowledge as an output affect the intentional direction of technological change and the emergence and decline of technological systems. In all cases, such price levels are determined by the idiosyncratic characteristics of the localized regional, historical, institutional and industrial context (Antonelli, 2009).

Firms have a clear incentive to search for potential complementarities between internal and external factors and intentionally characterize their innovative strategies so as to implement the interface between internal and external factors, achieve dynamic complementarities and increase their productivity and profitability. Let us analyze these two aspects in more detail.

Pecuniary knowledge externalities are not always and exclusively positive. Agglomeration in geographical and technological space, respectively within technological clusters and technological systems, has negative effects that are seldom identified. The density of firms accessing the same knowledge pools may have negative consequences in terms of reduced appropriability of technological knowledge. The clustering of firms in the same region favors the uncontrolled mobility of qualified workers and hence the leakage of sensitive information and competence. The likelihood of informal contacts among workers of different companies is increased and favored by repeated interactions and the complementarity and interdependence of research activities. Once more firms are exposed to the uncontrolled loss of proprietary knowledge.

As Kenneth Arrow (1962a) has pointed out, knowledge is indeed characterized by non-rivalry in its use. While two or more parties cannot share the simultaneous usage of the same tangible good, repeated usage of knowledge by many parties at the same time is possible. Each user does not deprive or limit the conditions of usage of other parties.

Knowledge however is characterized by substantial rivalry in exchange. Firms can extract substantial monopolistic rents from the exclusive command of original and unprecedented technological knowledge. The innovative firm can charge monopolistic prices for products that embody new technological knowledge as long as it is able to retain its exclusive control.

Non-rivalry in use and non-rivalry in exchange coincide only when perfect competition applies. But perfect competition applies only when all firms have access to all technological knowledge available with no restriction. When the access to knowledge is restricted, perfect competition no longer applies. Knowledge holders have a clear incentive to delay the dissemination and leakage of knowledge to third parties.

In Schumpeterian competition non-rivalry in use and non-rivalry in exchange differ widely. As it is well known in fact the exclusive control of proprietary technological knowledge impedes imitation and hence stretches the duration of monopolistic rents.

The access to the same pools of knowledge reduces the costs of external knowledge as an input into the generation of new knowledge, but reduces also its appropriability.

5. Pecuniary knowledge externalities at work

External knowledge is a non-disposable input because nobody can command all the knowledge available at any point in time. Internal and external knowledge are complementary inputs that it is necessary to combine in order to produce new technological knowledge. Following Nelson (1982) and Antonelli (1999) we can specify a knowledge production function.

In our case, the production and costs functions of knowledge can be stylized as follows:

$$T = (IK^a EK^b) \text{ with } a + b = 1 \tag{7.1}$$

$$C = pIK + uEK \tag{7.2}$$

where T represents new technological knowledge generated with constant returns to scale by means of internal knowledge (IK) and external knowledge (EK). Here p and u represent their respective unit costs.

The unit costs of internal knowledge consist of the market price for the resources that is necessary to enable the performance of research and development activities. The costs of external knowledge consist of the resources that are necessary to enable screening, understanding, purchasing and acquiring knowledge possessed by other agents in the system, including non-trivial efforts in terms of knowledge communication with regard to reception and absorption activities and knowledge networking. Such technological knowledge does not spill freely into the air. Dedicated activities are necessary in order to identify and acquire it. Moreover additional resources are necessary in order to make a new use of it. The acquisition of external knowledge is not free: in fact pecuniary externalities apply instead of technological externalities.

There are conducive contexts characterized by high quality knowledge governance mechanism in which, because of knowledge-non-exhaustibility, the costs of reproduction of technological knowledge are far below the costs of generation. Because of pecuniary knowledge externalities, the costs of external knowledge (u) are lower than the costs of internal sources of new knowledge (p) and below equilibrium levels (u^*). The latter would hold if and when knowledge was a normal economic good (see Figure 7.1).

Pecuniary knowledge externalities are found within economic systems where the costs of external knowledge are below equilibrium levels. Pecuniary knowledge externalities are found when and where knowledge reproduction costs differ sharply from generation costs, and knowledge governance at the system levels is effective and the efficiency of knowledge governance mechanisms is high. It is important to enphasize again here how important are knowledge governance costs. When knowledge governance costs are high, the actual costs of external knowledge are close to 'equilibrium' levels. Hence there are no pecuniary knowledge externalities. When knowledge governance costs are high there are actual knowledge spillovers and new growth dynamics cannot take place.

When pecuniary knowledge externalities apply, the maximizing firm will find the equilibrium in point B and produce a larger quantity of knowledge (T). The equilibrium technique will consist of a larger use of external knowledge with respect to internal knowledge. In a system characterized by positive pecuniary knowledge externalities, the firm will produce more technological knowledge than in a system where external knowledge has higher costs.

With positive pecuniary knowledge externalities in the upstream production of technological knowledge, the costs of technological knowledge generated by the firm are below equilibrium level: $s < s^*$.

This has important implications with respect to the output that the firm will produce. As it is shown in Figure 7.2, because of the upstream positive effects of external knowl-

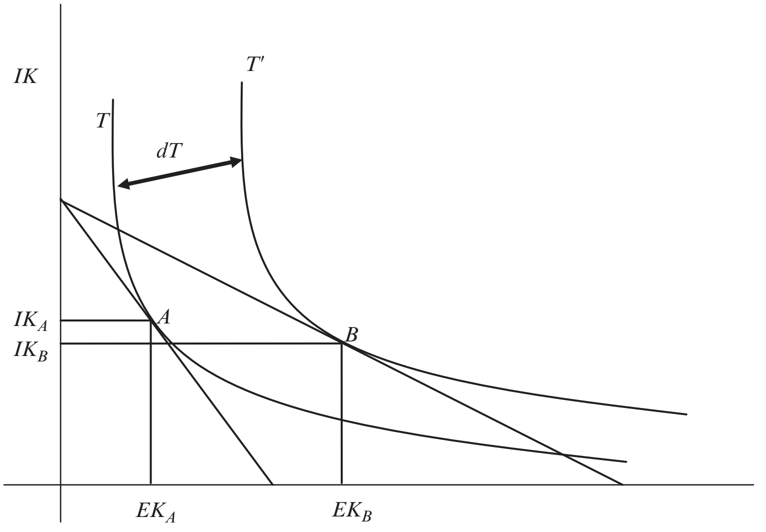


Figure 7.1 The Nelson knowledge production function

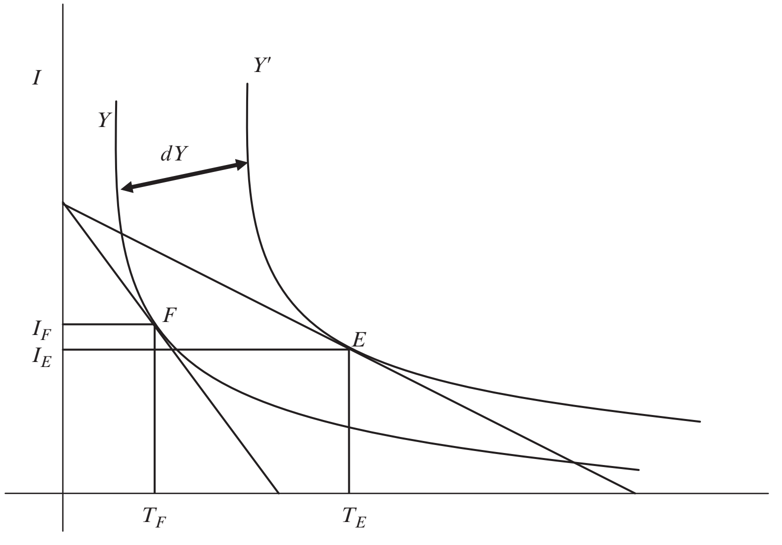


Figure 7.2 The Griliches production function

edge available at costs that are below equilibrium levels, the firm will be able to generate technological knowledge at lower costs and hence to produce a larger quantity of Y. The firm will select in fact the equilibrium point E, instead of F, where the firm that has no access to pecuniary knowledge externalities would go. The equilibrium in E implies a smaller demand for the bundle of tangible inputs (I), a more intensive use of the technology (T) and a larger output Y. As a matter of fact the amount of excess output dY generated by the firm that can take advantage of positive pecuniary knowledge externalities can be considered the residual, that is the excess output that cannot be explained in equilibrium conditions.

Following Griliches (1979), technological knowledge enters directly a standard Cobb-Douglas production function with constant returns to scale:

$$Y = (I^f T^g) \qquad \text{where } f + g = 1 \tag{7.3}$$

$$C = cI + sT \tag{7.4}$$

$$dY/Y = A \tag{7.5}$$

where for the sake of simplicity I is a bundle of tangible inputs, c are their costs, T is technological knowledge and s its cost, and A measures total factor productivity growth stemming from pecuniary knowledge externalities.

Total factor productivity growth can be explained by means of positive pecuniary knowledge externalities, because knowledge is a production factor both for the production of goods and for the generation of further knowledge, and it is characterized by non-exhaustibility and its production function is shaped by the complementarity between external and internal sources knowledge.

The working of pecuniary knowledge externalities is compatible with equilibrium conditions at the firm level while at the aggregate level the system is far from equilibrium. As long as pecuniary knowledge externalities are found, the typical system dynamics, stemming from the positive feedback generated by knowledge non-exhaustibility and knowledge complementarity, implemented by good knowledge governance mechanisms, are at work at the system level.

The application of the notion of pecuniary externalities to the economics of knowledge makes it possible to appreciate and identify both the positive and negative effects of knowledge appropriability. The tradition of analysis based on the contributions of Arrow (1962a) has focused the negative effects of the non-appropriability of knowledge in terms of missing incentives to the generation of new knowledge. The tradition of analysis paved by Griliches has appreciated the positive effects of the non-appropriability of knowledge in terms of the uncontrolled spillover of knowledge from 'inventors' to third parties. The two approaches however have rarely been integrated in a single framework. This is possible when the dual role of knowledge as both the output of an intentional economic process and an input in the production of both new knowledge and other goods that use knowledge as an intermediary factor is appreciated (Antonelli, 2008b).

The understanding of the dual role of knowledge as an output and an input makes it possible to use the notion of derived demand. The non-appropriability of knowledge, as sketched by the dotted supply and demand schedules of Figure 7.3, has effects both on demand and on supply. The effects on the derived demand side are negative, as the prices for knowledge are lower than they would be with a pure private good characterized by full appropriability and exhaustibility, and push downward the position of the demand curve (see in Figure 7.3 the shift of the derived demand curve from the bold to the dotted line), but the effects on the supply side are positive as they push downward the supply curve as well. The costs of the production of new knowledge are in fact lower – with

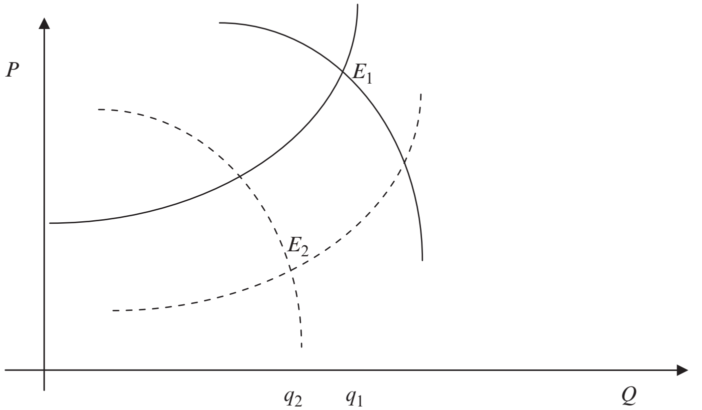


Figure 7.3 Positive and negative pecuniary knowledge externalities

respect to pure private goods – when the access to external knowledge that spills from original inventors is possible at a cost than is lower than in the case of a pure private good. Hence the actual amount of knowledge generated in the system may be much closer to equilibrium levels. The prices of knowledge may be lower. The actual extent to which such positive and negative effects apply and their relative weight depend very much upon the localized conditions of action of agents.[4]

The characteristics of the system in terms of knowledge governance mechanisms and hence the levels of knowledge transaction, communication and interaction costs are crucial to assess the long-term viability of the system dynamics. The analysis so far has not taken into account the negative effects of the number of agents that are active in the same knowledge pool and share basic knowledge complementarities on the levels of knowledge appropriability and hence the price for the products that embody new technological knowledge.

When such effects are taken into account we can introduce the notion of net pecuniary knowledge externalities. Net pecuniary knowledge externalities are the result of the joint assessment of both negative and positive effects. The latter reduce the costs of external knowledge and consequently, via the increase in total factor productivity, the costs of goods. The former reduce the price at which either knowledge as a good or as an input that incorporates the new knowledge can be sold. The levels of net pecuniary externalities depend on the combined effects of positive and negative pecuniary knowledge.

It is easy to derive the formal conditions to identify the optimum size of the local pools of knowledge. We know that the value of output depends on the quantity and its price P:

$$Y = PQ \tag{7.6}$$

The price at which the good that embodies new knowledge is sold is influenced by the number N of firms that have access to the same knowledge pool:

$$P = P^* / z(N) \tag{7.7}$$

We know that the cost of the output is influenced by the costs of external knowledge, which in turn are affected by the same number of firms N:

$$C = C^* / v\,(N) \tag{7.8}$$

$$dY/dN = dC/dN \,/\, dp/dN \tag{7.9}$$

The specific form of interplay between the positive effects on the costs of external knowledge and the negative effects on knowledge appropriability can acquire a quadratic form. In such circumstances the dynamics of the process will follow an S-shaped path. Formally we see that the dynamics of net positive externalities as dependent on the number N of firms may take a quadratic shape:

$$dN(t) = n\,(N(t) - (N^2(t)) \tag{7.10}$$

In fact equation (7.10) admits the standard logistic equation as a solution (see Figure 7.4).

Net pecuniary knowledge externalities provide the incentive to enter the knowledge pools. Entry will take place as long as they are positive. The flow of entry will take a quadratic shape and accelerated flows of entry are likely to take place in the proximity of the optimum size of the cluster. Beyond that level, firms will enter at a reduced pace. Entry will stop as soon as the negative effects of reduced knowledge appropriability are larger than the positive effects in terms of reduced costs of external knowledge.

It is clear in fact that agglomeration is no longer an unconstrained recipe. The size of the local knowledge pools can be too little or too large. There is also, according to specific conditions, an 'optimum' size of the local pools of knowledge.

6. The emergence and decline of knowledge systems

The availability of external knowledge explains, at one and the same time, the emergence of path-dependent local and sectoral systems of innovations, the features of regional specialization of regions, and the directionality of technological change. The availability of external knowledge in fact pushes firms to increase the complementarity of their knowledge base so as to take advantage of emerging pecuniary externalities in knowledge generation.

Technological knowledge cannot be regarded as an undifferentiated homogeneous body but rather as a bundle of highly idiosyncratic and circumscribed items possessed by individual agents, localized by their own circumscribed competences, more or less interconnected by a web of communication channels. At each point in time innovation systems are the nodes of such networks of communication channels. They are the result of the intentional search and implementation of knowledge complementarities among agents originally dispersed in knowledge space or clustered in other nodes. Learning agents, able to generate new knowledge, try to exploit pecuniary externalities of knowledge and move within the knowledge space, hence they direct technological knowledge towards well-specified characteristics that are shaped collectively. The collective directionality of technological knowledge that is generated through the convergence contributes to the emergence of structured collective pools of knowledge. Such innova-

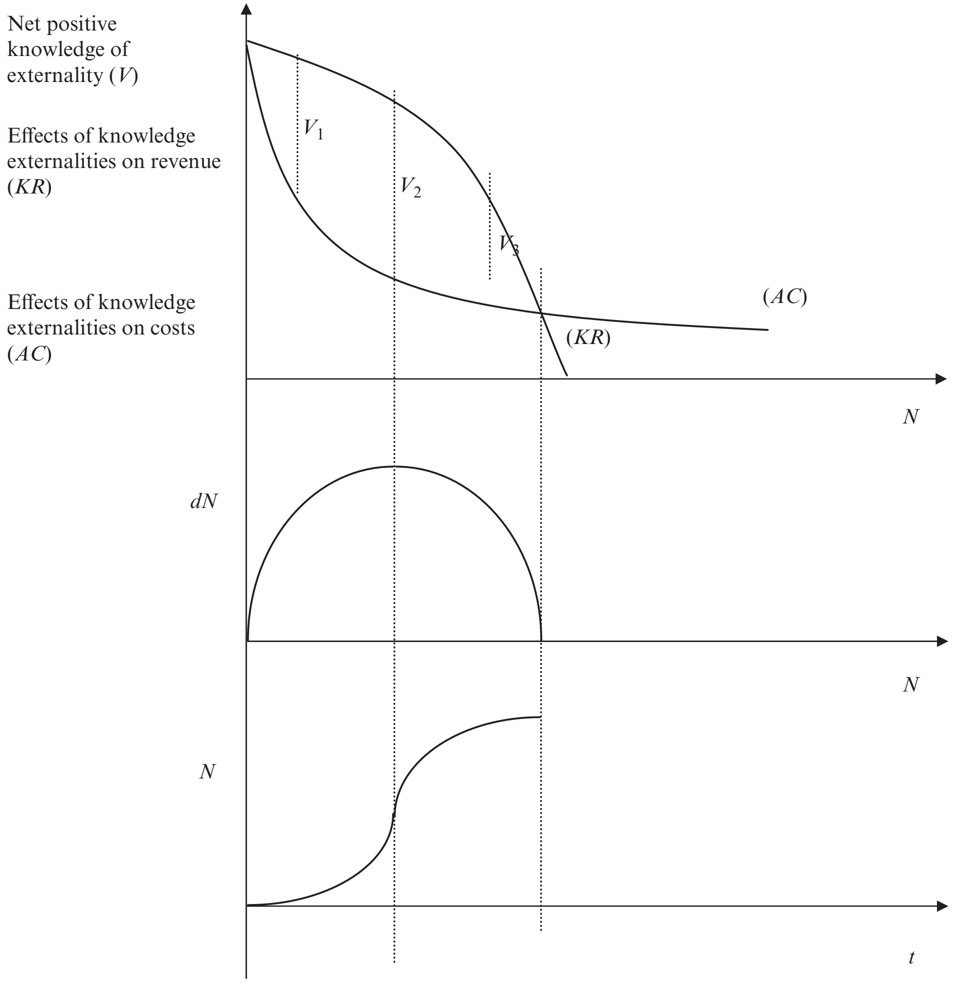


Figure 7.4 *The dynamics of the pecuniary knowledge externalities trade-off and the S-shaped diffusion process of innovation systems*

tion systems, characterized by the structure of communication channels and knowledge complementarities, exhibit dynamic features (Durlauf, 2005; Frenken, 2006; Guiso and Schivardi, 2007).

An understanding of the effects of pecuniary knowledge externalities in shaping the direction of technological change makes it possible to explain, at one and the same time, the convergence of the technological paths of firms towards local pools of complementary knowledge and the emergence of structured systems of innovation based on the local availability of distinctive sources of both tangible and intangible inputs (Durlauf and Johnson, 1992).

In this light, local and sectoral systems of innovation are no longer the exogenous result of fortuitous, institutional blending (Boschma, 2005; Malerba, 2005). They can be seen as nodes of communication channels that are the result of an endogenous emerging process that shares the complex dynamics of the creation of the Internet network (Antonelli, 2007; Pastor –Satorras and Vespignani, 2004).

Take as a starting point a population of heterogeneous and idiosyncratic firms, distributed in different regions, that have access to different sources of knowledge and factor markets and have a well-defined location in knowledge space. Each firm, exposed to a mismatch between beliefs and related plans, and the eventual factor and market conditions, is pushed to generate new knowledge and to introduce new technologies. This creative reaction requires dedicated activities: valorizing internal learning, carrying out formal research and development activities, and acquiring external knowledge, both tacit and codified. Such activities entail the assessment of specific costs, such as the costs of the coordination of the valorization of internal learning, the knowledge transaction costs necessary to purchase codified knowledge in the markets for knowledge, and the networking costs necessary to implement the acquisition of external knowledge, both codified and tacit. Even tacit external knowledge does not spill freely into the air: its acquisition is itself the result of intentional activities. Relevant absorption costs add to the actual purchasing costs of external knowledge. The selection of the kind of technological knowledge is affected both by the conditions for its generation and by the conditions for its exploitation. Each firm has a clear incentive to direct the generation of new technological knowledge towards applications that make it possible to combine its internal competence with the knowledge inputs that are locally abundant and that have a strong idiosyncratic character. The intentional technological convergence, in fact, makes it possible to generate more knowledge and to produce it at a lower cost.

Each firm discovers that the convergent alignment of its internal research activities with the complementary research activities of other firms, co-localized in both geographical and knowledge space, is a powerful factor of competitive strength. It is immediately clear that the lower the unit costs of external knowledge, the larger the amount of knowledge that the firm is able to generate, and the wider its localization in the specific context will be. A firm that is located in a conducive knowledge environment, and that is able to identify and access the local pools of knowledge at low costs, is induced to take advantage of it and hence to root the generation of its new knowledge in the characteristics of the environment into which it is based.

The amount of knowledge generated by firms is larger when they are able to align their research strategies in such a way as to take advantage of locally abundant knowledge. Consistently, in the downstream applications, firms can rely on a larger increase in efficiency with the same amount of budget available to fund the generation of new knowledge. The amount of external knowledge that is used in the knowledge generation process has a direct bearing not only on the amount of knowledge being generated and hence on the shift efficiency engendered in the production process, but also on its characteristics. Firms that rely more on external knowledge are more likely to produce complementary knowledge. The technological path of each firm will reflect the characteristics of both its own internal quasi-irreversibilities and learning processes, and the local context. The initial conditions play a key role in defining the context of action. The external context however, at each point in time, affects the dynamics powerfully. The direction of the process is constrained by the initial conditions, but it can change at each point in time. The path dependence limits, at each point in time, the span of possible directions. Path dependence consists of the continual redefinition of a limited span of possible directions. The convergence of the research strategies of each firm can gain momentum.

There are strong incentives that favor the convergence of the research projects of each firm towards local pools of common knowledge. Positive feedbacks are likely to reinforce the process as the efforts to increase the complementarity of the research activities of firms create the local pools of knowledge, with the result that the chances to access external knowledge increase. At the population level, the effects of individual convergence are reinforced by selection mechanisms. The success of the localized knowledge generation strategies acts as a powerful focusing mechanism that, by means of selection processes, favors the survival and growth of firms that have selected a convergent path of knowledge generation.

Systems of innovation emerge. They develop in technological districts and clusters, when the generation of new technological knowledge is reinforced by the emerging structure of complementarities implemented by communication channels provided by the intentional research strategies of firms that discover new sources of complementarities and move in the knowledge space. In special circumstances the emergence of innovation systems driven by highly performing network structures that have emerged through the collective dynamics of a myriad of agents in search for potential complementarities may lead to Schumpeterian gales of innovations.

At each point in time each firm is rooted in a well-defined location in knowledge space. The dynamics of the emergence (and decline) of technological systems based on potential knowledge complementarities can be grasped when we appreciate that each firm is able to move in such a knowledge space and generate new knowledge taking advantage of increased proximity and reinforced communication channels with other firms clustering in nodes. As a result, new systems of innovation based on nodes of coherent knowledge complementarity emerge (and others decay) while the direction of technological knowledge is shaped by the emergent collective convergence of each firm's research strategy (David et al., 1995).

Pecuniary externalities however are neither exogenous nor, by definition, static. The convergent dynamics may exhibit both positive and negative effects. On the one hand the amount of external knowledge available within the district keeps increasing and its costs are lower and lower. On the other hand, however, knowledge governance costs may increase as the number of firms accessing the same knowledge pools increase because of the effects of congestion in coordination. Density may have negative effects in terms of reduced knowledge appropriability: the case for excess clustering can take place when proximity favors the uncontrolled leakage of proprietary knowledge. In the same way the price of idiosyncratic inputs may increase with the increasing levels of their derived demand as shaped by the introduction of directional technological change.

The dynamics of the process reflect the interplay between the positive and negative changes of the levels of pecuniary externalities, both in knowledge generation and knowledge exploitation. The convergence of the direction of technological change and the emergence of innovation systems in geographical and technological space takes place as long as the rising levels of knowledge governance costs, and the rising prices of the idiosyncratic inputs do not exhaust pecuniary externalities. Innovation systems emerge depending on the relative weights of the positive and negative pecuniary externalities.[5] Innovation systems can emerge and decline depending on specific factors such as the characteristics of the technological knowledge, and the types of competition in product and factor markets, and the institutional context (Beaudry and Breschi, 2003).

The characteristics of the system in terms of knowledge governance mechanisms and hence the levels of knowledge transaction, communication and interaction costs are crucial to assess the long-term viability of the system dynamics.

7. Conclusions and policy implications

Technological and scientific knowledge is a collective, highly imperfect and heterogeneous activity. First of all it is not only an output, but also an input, an essential intermediary factor of production that is relevant both in the generation of new technological knowledge and in the generation of other goods. The dynamic efficiency of each firm and of the system at large depends on the factors affecting the generation and dissemination of knowledge.

The identification of the dual characteristics of technological knowledge as both an output and an input in the production of other goods and in the production of further knowledge, together with the understanding of the intrinsic complementarity between external and internal sources of knowledge, both non-disposable inputs in the generation of new knowledge, make it possible to apply the notion of pecuniary externalities in a novel context.

Pecuniary knowledge externalities are a powerful analytical tool that applies to the analysis of external knowledge as a necessary and yet costly production factor in the generation of new knowledge. The use of the notion of 'technological' externalities is consistent with the view that external knowledge falls from heaven like manna and spills freely into the atmosphere.

Pecuniary knowledge externalities are not always and universally positive. Agglomeration within geographic and technological clusters can yield negative consequences in terms of reduced appropriability of proprietary knowledge. Agglomeration within clusters yields positive effects only when the effects of pecuniary knowledge externalities on the costs of external knowledge are stronger than the effects of pecuniary knowledge externalities on the prices of the goods that embody new technological knowledge. A clear case for excess agglomeration has been identified in terms of reduced knowledge appropriability. Uncontrolled leakage and reduced exclusivity of proprietary knowledge can impede the long-term sustainability of such a process of self-propelling growth.

The quality of knowledge governance mechanisms that include the assessment of intellectual property right regimes is crucial for the actual viability of public policies based on knowledge externalities.

Such results call attention to the role of a public knowledge policy. The need for an economic policy regarding the production and dissemination of knowledge seems stronger than ever. Spontaneous knowledge governance mechanisms need to be complemented by a public policy. The implementation of the institutional set-up by means of policy actions that reduce uncertainty and create information, so as to reduce the effects of bounded rationality and information loads, seems to be a viable strategy to reduce the divide between profit maximization and social welfare. Public policy can reduce the major limits of the knowledge governance system so as to favor a more effective system of producing and circulating knowledge, with interventions aimed at increasing the amount of information each agent has access to.

Public knowledge policies can play a key role in encouraging dynamic coordination among the variety of heterogeneous players involved in the generation of knowledge as a complex and collective process. The state can favor the activity of interface bodies that have the specific mission to increase the dissemination of scientific knowledge and its communication to potential users. The creation of such interface agencies can increase the efficiency of the workings of the knowledge governance systems. Public interface agencies can help to identify the supply buried in the stocks of knowledge, often in the public domain, in universities and other public research centers, and awaken demand for its application. The role of public interface agencies is to push the academic community towards the market place and selected segments of the business community towards the academic one. Small firms are not even present in the knowledge markets. The minimum threshold of performance or research activity is often beyond the size possible for single small companies.

Moreover the state can specialize in the direct supply of knowledge, by means of university and public research centers, especially when it has high levels of fungibility, that is to say, knowledge with a wide range of applications in a broad array of activities and high levels of incremental enrichment. Public implementation of the access conditions to such knowledge, viewed as an essential facility, is the key to dynamic efficiency in the generation of new knowledge.

Notes

1. I acknowledge the funding of the European Union Directorate for Research, within the context of the Integrated Project EURODITE (Regional Trajectories to the Knowledge Economy: A Dynamic Model) Contract No. 006187 (CIT3), in progress at the Fondazione Rosselli and BRICK at the Collegio Carlo Alberto. I am grateful for the comments of many colleagues at the presentations of preliminary version at the European Science Foundation Exploratory Workshop 'Evolutionary Economic Geography' held at St Catharine's College, University of Cambridge, 3–5 April 2006; Conference 'Information, Intellectual Property, and Economic Welfare' held at the Fondazione Einaudi, Torino, 15–16 May 2006; Conference 'Déterminants et impacts économiques du management des connaissances', 23 November 2006, organized by the Ecole Superieure de Commerce et Management of the Université Catholique de Lyon and by the LEFI (Laboratoire d'Economie de la Firme et des Institutions) de l'Université Lyon Lumiere 2; Conference 'Libertad y Crescimiento' Madrid, 29–30 March 2007; Workshop 'Regional technological trajectories. Theoretical backgrounds and empirical observations' organized by the Max Planck Institute Evolutionary Economics Group, Jena, Germany, 27–29 September 2007. A preliminary version circulated as the No. 0709 of the Working Papers on Economics and Evolution of the Max Planck Institute of Economics, Evolutionary Economics Group, Jena. Last but not least, I acknowledge the useful comments of the Editors and other anonymous referees.
2. Past dependence defines dynamic processes characterized by high levels of sensitivity of the initial conditions. Path dependence emphasizes the possibility of changing the direction of non-ergodic processes.
3. As Scitovsky (1954) notes: 'This latter type of interdependence may be called pecuniary external economies to distinguish it from technological external economies of direct interdependence (p. 146).
4. At the firm level it is clear that firms have a strong incentive to implement, on the one hand, exploitation strategies in order to reduce the negative effects of non-appropriability and hence minimize the uncontrolled leakage of their proprietary knowledge and, on the other, exploration strategies, in order to maximize the benefits of knowledge spilling into the atmosphere. It is clear that both the exploitation and exploration of knowledge require dedicated activities and relevant resources.
5. In specific contexts the interplay can lead to logistic processes of emergence with S-shaped dynamic process that identify critical masses. See Antonelli (2007).

Bibliography

Antonelli, C. (1995), *The Economics of Localized Technological Change and Industrial Dynamics*, Boston, MA: Kluwer Academic Publishers.

Antonelli, C. (1999), *The Microdynamics of Technological Change*, London: Routledge.

Antonelli, C. (2001), *The Microeconomics of Technological Systems*, Oxford: Oxford University Press.
Antonelli, C. (2007), 'The system dynamics of collective knowledge: from gradualism and saltationism to punctuated change', *Journal of Economic Behavior and Organization*, **62**, 215–36.
Antonelli, C. (2008a), *Localized Technological Change: Towards the Economics of Complexity*, London: Routledge.
Antonelli, C. (2008b), 'Pecuniary knowledge externalities: the convergence of directed technological change and the emergence of innovation systems', *Industrial and Corporate Change*, **17**, 1049–70.
Antonelli, C. (2009), 'Localized appropriability: pecuniary externalities in knowledge exploitation', *Technology Analysis and Strategic Management*, **21**, 727–42.
Antonelli, C., P.P. Patrucco and F. Quatraro (2008), 'The governance of localized knowledge externalities', *International Review of Applied Economics*, **22**, 479–98.
Arrow, K.J. (1962a), 'Economic welfare and the allocation of resources for invention', in R.R. Nelson (ed.), *The Rate and Direction of Inventive Activity: Economic and Social Factors*, Princeton, NJ: Princeton University Press for NBER, pp. 609–25.
Arrow, K.J. (1962b), 'The economic implications of learning by doing', *Review of Economic Studies*, **29**, 155–73.
Arrow, K.J. (1969), 'Classificatory notes on the production and transmission of technical knowledge', *American Economic Review*, **59**, 29–35.
Atkinson, A.B. and J.E. Stiglitz (1969), 'A new view of technological change', *Economic Journal*, **79**, 573–8.
Beaudry, C. and S. Breschi (2003), 'Are firms in clusters really more innovative?', *Economics of Innovation and New Technology*, **12**, 325–42.
Boschma, R.A. (2005), 'Proximity and innovation: a critical assessment', *Regional Studies*, **39**, 61–74.
Boschma, R.A. and K. Frenken (2006), 'Why is economic geography not an evolutionary science? Towards evolutionary economic geography', *Journal of Economic Geography*, **6**, 273–302.
Boschma, R.A. and R. Martin (2007), 'Editorial: Constructing an evolutionary economic geography', *Journal of Economic Geography*, **7**, 537–48.
Breschi, S. and F. Malerba (eds) (2005), *Clusters Networks and Innovation*, Oxford: Oxford University Press.
Cohen, W.M. and D.A. Levinthal (1989), 'Innovation and learning: the two faces of R&D', *Economic Journal*, **99**, 569–96.
Cohen, W.M. and D.A. Levinthal (1990), 'Absorptive capacity: a new perspective on learning and innovation', *Administrative Science Quarterly*, **35**, 128–52.
David, P.A. (1993), 'Knowledge property and the system dynamics of technological change', *Proceedings of the World Bank Annual Conference on Development Economics*, Washington, DC: The World Bank.
David, P.A. (1994), 'Why are institutions the "carriers of history"? Path dependence and the evolution of conventions, organizations and institutions', *Structural Change and Economic Dynamics*, **5**, 205–20.
David, P.A., D. Foray and J.M. Dalle (1995), 'Marshallian externalities and the emergence and spatial stability of technological enclaves', *Economics of Innovation and New Technology*, **6**, 147–82.
David, P.A. and J.L. Rosenbloom (1990), 'Marshallian factor market externalities and the dynamics of industrial localization', *Journal of Urban Economics*, **28**, 349–70.
D'Ignazio, A. and E. Giovannetti (2006), 'From exogenous to endogenous economic networks: internet applications', *Journal of Economic Survey*, **20**, 757–96.
Durlauf, S.N. (2005), 'Complexity and empirical economics', *Economic Journal*, **115**, 225–43.
Durlauf, S. and P.A. Johnson (1992), 'Local versus global convergence across national economies', Working Paper No. 3996 of the NBER, Cambridge.
Feldman, M.P. (1999), 'The new economics of innovation spillovers and agglomeration: a review of empirical studies', *Economics of Innovation and New Technology*, **8**, 5–26.
Foss, N. (1997), *Resources Firms and Strategies. A Reader in the Resource-Based Perspective*, Oxford: Oxford University Press.
Foss, N.J. (1998), 'The resource-based perspective: an assessment and diagnosis of problems', *Scandinavian Journal of Management*, **15**, 1–15.
Foss, N. and V. Mahnke, (eds) (2000), *Competence, Governance and Entrepreneurship. Advances in Economic Strategy Research*, Oxford: Oxford University Press.
Freeman, C. (1991), 'Networks of innovators: a synthesis of research issues', *Research Policy*, **20**, 499–514.
Frenken, K. (2006), 'Technological innovation and complexity theory', *Economics of Innovation and New Technology*, **15**, 137–55.
Frenken, K., F. Van Oort, T. Verburg, (2007), 'Related variety, unrelated variety and regional economic growth', *Regional Studies*, **41**, 685–97.
Griffith, R., S. Redding, J. Van Reenen, (2003), 'R&D and absorptive capacity: theory and empirical evidence', *Scandinavian Journal of Economics*, **105**, 99–118.
Griliches, Z. (1979), 'Issues in assessing the contribution of research and development to productivity growth', *Bell Journal of Economics*, **10**, 92–116.

Griliches, Z. (1992), 'The search for R&D spillovers', *Scandinavian Journal of Economics*, **94**, 29–47.

Grossman, G.M. and E. Helpman, (1994), 'Endogenous innovation in the theory of growth', *Journal of Economic Perspectives*, **8**, 23–44.

Guiso, L. and F. Schivardi, (2007), 'Spillovers in industrial districts', *Economic Journal*, **117**, 68–93.

Hayek, F.A. (1945), 'The use of knowledge in society', *American Economic Review*, **35**, 519–30.

Jaffe, A.B. (1986), 'Technological opportunity and spillover of R&D: evidence from firms' patents, profits and market value', *American Economic Review*, **79**, 985–1001.

Jones, C.I. (2002), *Introduction to Economic Growth*, 2nd edn, New York: Norton.

Kaldor, N. (1981), 'The role of increasing returns technical progress and cumulative causation', *Economie Appliquée*, **34**, 593–617.

Malerba, F. (2005), 'Sectoral systems of innovation: a framework for linking innovation to the knowledge base, structure and dynamics of sectors', *Economics of Innovation and New Technology*, **14**, 63–82.

Mansfield, E., M. Schwartz and S. Wagner, (1981), 'Imitation costs and patents: an empirical study' *Economic Journal*, **91**, 907–18.

Marshall, A. (1890), *Principles of Economics*, 8th edn (1920), London: Macmillan.

Martin, R.L. (1999), 'The new geographical turn in economics: some critical reflections', *Cambridge Journal of Economics*, **23**, 65–91.

Martin, R. (2007), 'The localization of industry', in T. Raffaelli, G. Becattini and M. Dardi (eds), *The Elgar Companion to Alfred Marshall*, Cheltenham, UK and Northampton, MA, USA: Edward Elgar.

Meade, J.E. (1952), 'External economies and diseconomies in a competitive situation', *Economic Journal*, **62**, 54–67.

Milgrom, P. and J. Roberts (1995), 'Complementarities and fit. Strategy structure and organizational change in manufacturing', *Journal of Accounting and Economics*, **19**, 179–208.

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

Nelson, R.R. (1982), 'The role of knowledge in R&D efficiency', *Quarterly Journal of Economics*, **97**, 453–70.

Nelson, R.R. (ed.) (1993), *National Systems of Innovation*, Oxford: Oxford University Press.

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

Nooteboom, B. (2000), *Learning and Innovation in Organizations and Economics*, Oxford: Oxford University Press.

Nooteboom, B. (2003), 'Managing exploration and exploitation', in S. Rizzello (ed.), *Cognitive Developments in Economics*, London: Routledge.

Nooteboom, B. (2004), 'Governance and competence: how can they be combined?', *Cambridge Journal of Economics*, **28**, 505–25.

North, D.C. (1997), 'Some fundamental puzzles in economic history', in W.B. Arthur, S.N. Durlauf and D. Lane (eds), *The Economy as an Evolving Complex System II*, Redwood City, CA: Addison-Wesley.

Pastor-Satorras, R. and A. Vespignani (2004), *Evolution and Structure of the Internet*, Cambridge: Cambridge University Press.

Patel, P. and K. Pavitt (1994), 'National innovation systems: why they are important and how they might be measured and compared', *Economics of Innovation and New Technology*, **3**, 77–95.

Patrucco, P.P. (2009), 'Collective knowledge production, costs and the dynamics of technological systems', *Economics of Innovation and New Technology*, **18**, 295–310.

Penrose, E.T. (1959), *The Theory of the Growth of the Firm*, Oxford: Basil Blackwell (reprinted 1995, Oxford: Oxford University Press).

Pisano, G.P. (1990), 'The R&D boundaries of the firm: an empirical analysis', *Administrative Science Quarterly*, **35**, 153–76.

Romer, P.M. (1986), 'Increasing returns and long-run economic growth', *Journal of Political Economy*, **94**, 1002–37.

Romer, P.M. (1990), 'Endogenous technological change', *Journal of Political Economy*, **98**, S71–102.

Romer, P.M. (1994), 'The origins of endogenous growth', *Journal of Economic Perspectives*, **8**, 3–22.

Scitovsky, T. (1954), 'Two concepts of external economies', *Journal of Political Economy*, **62**, 143–51.

Scott A.J. and M. Storper (2007), 'Regions globalization development', *Regional Studies*, **41**, S191–S205.

Viner, J. (1931), 'Costs curves and supply curve', *Zeitschrift für Nationalokonomie*, **3**, 23–46.

Williamson, O.E. (1996), *The Mechanisms of Governance*, New York: Oxford University Press.

Young, A.A. (1928), 'Increasing returns and economic progress', *Economic Journal*, **38**, 527–42.

8 The relationship between multinational firms and innovative clusters

Simona Iammarino and Philip McCann

1. Introduction

In the globalised economy, the geographical dispersion of economic and innovative activities is occurring on a massive scale. However, geographic dispersion does not lead to the wonderland of a 'borderless world', and the centripetal forces of geography are not rescinded by globalisation. The breathtaking speed of geographical dispersion has been accompanied by strong processes of spatial concentration, and much of the recent cross-border expansion of manufacturing and services production, as well as of innovative activities, has been concentrated in a handful of specialised industrial clusters and regional systems of innovation, in both advanced and emerging economies.

Globalisation often occurs as an extension of national clusters across national borders. The first implication is that, while some stages of the value chain are internationally dispersed, others remain spatially concentrated. The second implication is that those internationally dispersed activities, and especially innovative activities, typically agglomerate in a limited number of overseas sub-national regions. This clearly indicates that agglomeration economies continue to matter. What needs to be better understood though, is how they are changing under the impact of globalisation or, in other words, what kind of challenges are posed by the new interaction of 'global versus local'. One crucial aspect of such a relationship deserves particular attention: the creation and diffusion of innovation, which, more than other economic processes, show rather complex patterns of spatial distribution.

Industrial clustering has been seen as a central component of the processes of regional innovation and growth. However, the observed variety of geographical innovative models has been often explained neglecting the nature of new knowledge in different production sectors and institutional settings. In particular, technological regimes, industrial structures and organisational practices, as well as their dynamics, are often overlooked in favour of simplified and stylised constructs on successful regions or clusters.

In this chapter our aim is threefold: (1) to present a typology of clusters that accounts for both the diversity and the dynamics of spatial configurations; (2) to explore how the innovative strategies of multinational enterprises (MNEs) and the different kinds of knowledge opportunity offered by clusters influence MNEs' location; and (3) to provide some illustrations of cluster transition from one spatial model to another, with consequent changes in MNE contributions to spillovers. In all these respects, an evolutionary approach is needed to overcome the static accounts of the interaction 'global versus local', and the insufficient consideration still devoted to both the nature of innovation processes and the structural conditions under which technical change occurs within firms and across space.

The chapter is divided into seven sections. In the following section we briefly discuss the ongoing changes underlying the locational choices of multinational enterprises' operations, with particular attention to innovative and technological activities. In section 3 we summarise some arguments regarding the relationship between geography and knowledge spillovers, and we present a transaction costs classification of spatial typologies previously developed elsewhere. Section 4 addresses the limitations of transaction cost views in analysing the processes of evolution of spatial models and the relationship between local contexts and global actors, by briefly introducing evolutionary perspectives on technical and structural change. Such perspectives are adopted in section 5 to extend and integrate the classification proposed in section 3, in order to give an account of the diversity and multiplicity of possible evolutionary paths of spatial configurations. Section 6 then brings in some emerging features of the interaction between MNEs and localised innovative economies, also using a few empirical examples. Section 7 outlines some conclusions and challenges for future research.

2. The geography of MNE innovative activities: 'concentrated dispersion'

The question of how globalisation affects the geographic dispersion of knowledge and innovation, and whether this fosters or constrains local technological capabilities, has been extensively discussed in theoretical debates on the role of foreign direct investments and MNEs (e.g. Dunning, 1993). More recently, this question has also received attention in innovation theory, the theory of the firm, and economic geography (e.g. Chandler et al., 1998). While traditional analyses of multinational activity emphasised the centralised nature of research and development efforts of multinational firms, contemporary contributions have argued the obsolescence of this thesis (e.g. Cantwell, 1995). The trend toward globalisation of technology and innovative activity has resulted in a questioning of the rather narrow role conferred on the MNE subsidiary by Vernon (1966), and in attempts to redefine MNE units as key creators of innovation and technological knowledge – a role originally suggested by Dunning (1970) and later developed by Chesnais (1988), Cantwell (1994) and Fors (1998) among others.

'Concentrated dispersion' has been used to indicate the ongoing reorganisation of MNEs' production by operating across borders through geographically dispersed and interdependent global networks, which show at the same time high spatial concentration at the sub-national scale (Ernst et al., 2001). Such a concentrated dispersion raises an important question: What factors explain why some value-chain activities are more prone to geographic dispersion, while others are more prone to be spatially clustered?

The usual suspects of course are differences in labour costs and knowledge-intensity. There is a strong presumption that high-wage and knowledge-intensive activities are more likely to be subject to agglomeration effects, and hence resistant to geographic dispersion. By the same token, geographic dispersion can be expected to be most prominent for low-wage and low-skill value-chain activities (Ernst et al., 2001). Much research on the evolving 'global versus local' interaction has thus been based on the distinction (industry-driven, not function-driven) between low-wage, low-skill declining industries, and high-wage, high-skill rising industries.

Such simple dichotomies, however, have often failed to provide convincing results, for two main reasons: low-wage, low-skill value stages are also found in most technology-intensive industries, and high-wage, high-skill activities exist in traditional industries; both competence requirements and industry boundaries keep changing over time, making an analytical focus on the industry level more problematic (Ernst et al., 2001). On the other hand, the labour cost argument seems more apt to explain the geographic dispersion of global value chains across macro-areas – and in particular the relocation of economic activities in general from advanced to emerging or less advanced economies – than to deal with spatial agglomeration phenomena occurring today in both industrialised and developing countries. Therefore, the rationale to explore the current patterns of clustering by MNEs' innovative and technological operations has shifted away from locational choices driven by the technology content dichotomy towards the organisational and institutional logic underlying both the firm and the host location.

Concentrated dispersion goes hand in hand with increasingly complex and systemic forms of integration through global networks. Knowledge flows and technology transfer occur as a consequence of new organisational modes within and outside the MNE, depending on the characteristics of different localised institutional settings involved in the network. Even firms that are labelled on the basis of the same industry- or sector-specific features (such as high- or low-tech) may differ greatly in terms of technological base, use different sources of innovation, and display specific patterns of interactions with local environments. This degree of complexity is all the more true when looking at multi-product and multi-technology firms, such as most contemporary large MNEs, that are likely to follow a variety of organisational modes and technological strategies across multiple locations.

The role of MNEs as creators of new technology (and not just as international transfer vehicles) has been facilitated by the trend for MNEs to establish internal and external networks for innovation. Internationally integrated networks within the firm may lead, through a greater focus in the specialisation of technological efforts in affiliates, to an improvement of innovation capacity both of the firm and of the host location. Inter-firm networks established between MNE affiliates and local firms may, in addition, amplify the advantages of geographical agglomeration in some particular lines of technological development. In other words, in order to renovate existing competencies, it is generally necessary for a firm to extend those capabilities into new related fields of production and technology, and across a variety of locations (Cantwell and Iammarino, 2003).

The firm is thereby able to benefit from the dynamic economies of scope that derive from the technological complementarities between related paths of innovation in spatially distinct institutional settings. In this perspective, MNEs spread the competence base of the firm and acquire new technological assets, or sources of technological advantage. Resources and capabilities that are critical for firms' competitive success 'can often be found inside a region, rather than within any single firm' (Enright, 1998, p. 315). For their part, indigenous firms may in principle benefit from localised knowledge spillovers from MNEs, given the access of the latter to complementary streams of knowledge being developed in other regions or clusters.

The precise form that knowledge networks take, and their effects, depend on many factors, among which are the knowledge base of the MNE, the nature of agglomeration forces that shape the spatial organisation of economic activities, the type of transactions and relationships between global and local firms, the intra-cluster absorptive capacity, and the extra-cluster linkages (Bell and Albu, 1999; Giuliani, 2005, and Chapter 12 in this book). Knowledge networks will be obviously affected by industry structures and sectoral compositions. Scale and scope economies in research and technological activities, which allow solving a high degree of technological complexity, are a prerogative of large firms, and in particular of MNEs. On the other hand, external economies are usually beneficial particularly to spur incremental innovative processes of small and medium sized enterprises (SMEs). The nature and extent of externalities cannot be separated from the dominant firm dimension in the cluster, and from the patterns of collaboration between small local firms, large indigenous firms and MNEs. Another issue to be considered is that MNEs can be at the same time 'internal' actors, contributing to the creation and diffusion of new technical knowledge within the cluster, and 'external' players, channelling knowledge created elsewhere (within the firm) into the local system, thereby assuming the role of technological gatekeepers, but even influencing the cluster knowledge base without being necessarily located *in situ* (as in the case of global buyers or distributors).

Even though the international business literature has largely acknowledged the geographical concentrated dispersion of production and technology networks as a chief mechanism to achieve sustainable competitive advantages in both MNEs and host regions (e.g. Almeida, 1997; Cantwell, 1989; Cantwell and Iammarino, 1998, 2003; Cantwell and Janne, 1999; Frost, 2001; Kogut and Chang, 1991; Kuemmerle, 1999; Pearce, 1999; Zander, 1999), the operational mechanisms of such technological linkages tend to be mainly depicted as a merely micro-to-micro (intra- and inter-firm) interaction, where the space itself is just an exogenous variable hardly explained in its internal structure and dynamics. As emphasised by Rugman and Verbeke (2002), the introduction of operational space into the study of MNE locational advantages is becoming essential. In other words, while the country level of analysis has been extended to (macro)regional trade and investment blocks (e.g. Ohmae, 1985; Rugman, 2005), as far as the (meso)subnational level is concerned (i.e. innovative clusters, regional innovation systems, local production systems, etc.) the reference literature is still rather scanty.

In the next section, we interpret the 'global versus local' interaction from the angle of the logic of spatial agglomeration by using a transaction costs-based perspective. This latter is especially useful to show that many stylised cluster notions found in the current literature are predicated on the basis of assumptions that are often incompatible with MNE behaviours.

3. Clustering, MNEs and knowledge spillovers: a transaction costs classification of spatial models

The most popular explanations of the benefits of spatial clustering have focused either on the role played by specialised external agglomeration economies (Marshall, 1920) – namely local information and knowledge spillovers, a local supply of non-traded inputs, and a skilled local labour pool – or on the role played by industrial variety in agglomeration, as in the models of the proponents of the new economic geography (e.g. Fujita et al., 1999; Krugman, 1991) and in traditional approaches to urbanisation and diversification economies (e.g. Glaeser et al., 1992; Jacobs, 1960). This literature suggests that inter-industry spillovers may be generally more important than intra-industry spillovers in explaining local growth (Martin and Ottaviano, 1999), although intra-industry effects may dominate in certain manufacturing activities (Henderson et al., 1995).

Nevertheless, the relationship between geography and economic growth appears to be rather more subtle, complex and varied than these agglomeration models suggest, particularly at a time when the inducements derived from global market competition and changes in demand, and from technological change, arise at an incredibly high speed. The original Marshallian-Arrow-Romer (MAR) stream of literature plus the variety literature on urbanisation and diversification economies have thus been broadened by including technological indices as endogenous variables in explaining local economic growth (e.g. Acs, 2002; Jaffe et al., 1993).

In the main, however, while these neoclassical interpretations have focused on localised spillover effects as the major analytical framework to explain cluster existence and growth, they have largely disregarded other possible mechanisms underlying spatial agglomerations (Autant-Bernard et al., 2003; Breschi and Lissoni, 2001; and Breschi et al., Chapter 16 in this book). In particular, production function-based models do not give a full account of the diversity of possible spatial configurations, nor do they tackle issues such as the evolution of clusters and regions and the disruptive changes imposed by technological progress and globalisation processes on agglomeration economies (Belussi and Sammarra, 2010; Hilpest, 2003).

As is mentioned in section 2, various factors are at work in shaping the relationships between MNEs and local environments, and therefore in determining the nature and extent of localised knowledge spillovers possibly stemming from such relationships. In this respect, it appears that we have two sets of fundamental questions to be addressed. First, we need to determine *whether* and *where* localised knowledge spillovers exist, and then second, we need to understand exactly *how* they do occur and change over time.

It has to be highlighted that the difficult analytical problems relating to the diverse features of spatial agglomerations are compounded by severe problems of identification and definition.[1] Here we provide a classification of spatial types that is independent of either the sector or the location, but instead is based on the microeconomic behaviour and objectives of the co-located actors, and on the transactions existing in the cluster or region.

A transactions costs approach was adopted elsewhere to present three different stylised sets of geography–firm–industry organisational relationships that exist in the literature (Gordon and McCann, 2000; McCann, 2001; McCann and Shefer, 2004; McCann and Sheppard, 2003; Simmie and Sennet, 1999). This classification is based on the (often implicit) assumptions underlying most of the existing literature on agglomeration phenomena. As such, the categories are not meant to be interpreted as representing any particular geographical place. These stylised characterisations of spatial clustering are distinguished in terms of the nature of firms in the cluster or region, and the nature of their relations and transactions undertaken within the specific local environment. They can be termed the *pure agglomeration*, the *industrial complex*, and the *social network*. The main characteristics of each of the spatial types, all quite different, are listed in Table 8.1.[2]

This classification is useful to illustrate that knowledge spillovers do not stem as an automatic mechanism from the location of MNEs in any particular spatial agglomeration, but, rather, they first depend on the nature and characteristics of the transactions taking place locally. In other words, this classification can help determine whether and where knowledge spillovers may occur.

From a transaction cost perspective, the firm's view of the net benefits of knowledge spillovers hinges on its assessment of the relative importance to itself of knowledge out-

Table 8.1 Industrial clusters: a transaction cost perspective

Characteristics	Pure agglomeration	Industrial complex	Social network
Firm size	Atomistic	Some firms are large	Variable
Characteristics of relations	Non-identifiable Fragmented Unstable frequent trading	Identifiable Stable and frequent trading	Trust Loyalty Joint lobbying Joint ventures Non-opportunistic
Membership	Open	Closed	Partially open
Access to cluster	Rental payments Location necessary	Internal investment Location necessary	History Experience Location necessary but not sufficient
Space outcomes	Rent appreciation	No effect on rents	Partial rental capitalisation
Example of cluster	Competitive urban economy	Steel or chemicals production complex	New industrial areas
Analytical approaches	Models of pure agglomeration	Location–production theory Input–output analysis	Social network theory (Granovetter)
Notion of space	Urban	Local or regional but not urban	Local or regional but not urban

flows and inflows, particularly when strong economies of scale and scope in research and technology are active within the firm (as in the case of large or multinational firms). As said above, the structure and organisation of an industry will affect the firm's perceptions of knowledge flows, and therefore, in particular, its choice to locate *in situ* technological and innovative operations.

For example, we can consider the case of a broadly competitive market structure that is characterised by a large number of firms, each with a relatively small market share and profits. In this case, competitor firms will have less to lose from knowledge outflows and more to gain from inflows stemming from a strong clustered location. In such a situation, the public good aspect of local knowledge will predominate and knowledge outflows will be viewed as being generally positive. In the taxonomy of Table 8.1, this combination of features is exhibited by the pure agglomeration model, where firms are essentially atomistic, their relations are inherently transient, and the notion of space is mostly urban, in that this type of clustering only exists within individual cities.

In an oligopolistic industry structure, on the other hand, particularly when characterised by a few large firms or MNEs, each with a large market share and considerable strategic interdependence, firms often perceive that knowledge outflows to industry rivals can be extremely costly in terms of lost competitive advantage. In these circumstances the private good aspect of knowledge is generally the dominant consideration. In situations where any knowledge outflows from a firm are more valuable to its competitors than are any potential knowledge inflows to the firm from its competitors, the overall net effect of unintended knowledge outflows will be perceived by the MNE to be negative

(McCann and Mudambi, 2004, 2005). This will lead the MNE to decide not to locate their technological operations in clusters characterised by pure agglomeration, although, depending on transportation and other types of cost, they may consider locating in complexes characterised by stable planned and long-term inter-firm relationships, and by a geographical configuration that is often local or regional but rarely urban.

This argument provides a powerful counter-argument to the Porter logic in favour of industrial clustering (Porter, 1990). This is because if we apply Akerlof's (1970) market-for-lemons model, regions and industrial clusters that include large oligopolistic competitors will be plagued by adverse selection and should either fail to form, or become concentrations of mediocrity, unless the clustering of firms exhibits the characteristics of an industrial complex in Table 8.1. This counter-argument to the Porter logic appears to explain the empirical finding that many of the largest MNEs do not co-locate their knowledge creation activities with those of their competitive rivals (Cantwell and Santangelo, 1999; Simmie, 1998).

It follows that the only possible reason why MNEs would locate their innovative activities in an environment characterised primarily by a pure agglomeration, would simply be as a means of facilitating the hiring of specialist labour. However, MNEs are often able to hire appropriate specialist labour or to tap into local technological expertise by simply locating within the same broad regional innovation system (Cantwell and Iammarino, 2003), rather than within the same urban location. According to this perspective, the inter-firm spillover arguments implicit in the pure agglomeration model are therefore not always applicable to oligopolistic MNEs (Arita and McCann, 2002a, 2002b).

Similar arguments are often also pertinent in the case of the social network model of Table 8.1, associated primarily with Granovetter's work (1973, 1985) and developed as a response to the hierarchical model of Williamson (1975). Such social networks are assumed to operate on the basis of trust relations. However, the notion itself of trust may vary greatly between social networks characterised by substantial relational symmetry, and those characterised by asymmetries among the clustered firms and centred on large leader-firms, either locally owned or foreign affiliates. This latter case can be assimilated to the 'hub-and-spoke' type of cluster identified by Markusen, where one or several firms/facilities in one or more sectors, usually large and vertically integrated (e.g. Boeing in Seattle or Toyota in Toyota City), act as anchors or hubs to the regional economy, with smaller suppliers and related activities spread around them like the spokes of a wheel, with different degrees of strength (Markusen, 1996). Interestingly, in no case will there be perfect symmetry among the various agents operating in the cluster, but one agent (or more of them) may play a distinct role, leading the cluster in terms of organisation, innovation, and/or finance (Guerrieri and Pietrobelli, 2004).

Thus, on the basis of inward and outward knowledge flows and appropriability arguments, the conditions under which such trust relationships will emerge naturally are largely opaque. If the firms in a cluster are to be small, then the arguments for pure agglomeration may be largely applicable in the case of the social network, although distinguishing between the two becomes problematic. However, if some of the firms in the cluster are to be large and/or MNEs, then it is not clear how such a social network would evolve, and much depends on the nature of transactions they establish with (large and small) local firms. The structure of vertical and horizontal linkages between MNEs and local firms and, within the first type of connections, the extent of backward and forward linkages, may be crucial in determining the occurrence and the extent of spillovers. Even in the same industry, clusters may be characterised by highly vertically integrated firms, or by stage production with significant sub-contracting linkages: the trust relationships and the role of reputation will differ considerably in such different spatial and industrial agglomerations (Guy, 2009).

Yet, the classification reported in Table 8.1 can only represent a partial and static picture. The actual detailed interplay between industries and geographical locations, or the specific evolutionary features of different regions and clusters, have been largely neglected by most current stylised economic geography approaches. This has caused hectic efforts to identify localised knowledge spillovers between local firms, MNEs and other organisations, irrespective of the characteristics of any functional dimension of knowledge processes depending on the type of industry and spatial agglomeration model.

4. The dynamics of spatial configurations: knowledge, and technological and structural change

One of the advantages of the transactions costs classification scheme reported in section 3 is that it provides an organising framework capable of dealing with the diversity of spatial concentrations we observe, offering possibilities for the inclusion of a technology and innovation component as an additional explanation for such a diversity. At the same time, a weakness of this framework is that, in the transaction cost perspective itself, hierarchies, and particularly, but not exclusively, firm structures, are reduced to a consequence of changes in transaction costs, whereas dynamic factors such as learning, accumulation, and knowledge creation are largely ignored. More specifically, the transaction cost approach underlying Table 8.1 has some limitations because of its essentially static nature, its narrow definition of knowledge and technology, and its discounting of the relationship between innovation processes and industry structures.

With respect to the first point, the classification presented in the previous section is mainly static, and reflects the bulk of the existing literature on the cluster concept that it summarises. Underlying most of such a literature there are once-and-for-all efficiency gains (in terms of economies of scale and scope, transaction and transport costs and input–output linkages). However, the relationship between firm location and technology is mainly dynamic (see among others Audretsch, 1997; Dosi, 1988; Dosi et al., 1997; Nelson, 1991; see also Bottazzi and Dindo, Chapter 24 in this book), both in terms of industrial demography (firm entry, exit, growth, relocation, etc.) and in terms of cluster life cycle (cluster birth, growth, decline, openness, attractiveness, etc.). Thus, clusters and regional systems are not necessarily static in that they may evolve over time, possibly blending the features of the three typologies into different combinations, and possibly shifting from one main type to another according to the relative stage in their life cycle. Dynamic agglomeration economies and the central role of learning and new knowledge creation therefore lead to distinctive paths of regional evolution.

Regarding the second point, the logic of the contractual transaction cost arguments about appropriability of the returns on innovation, widely applied to the perception of knowledge inflows and outflows by MNEs, depends to a large extent on a narrow definition of technology. The traditional approach in economics has regarded knowledge as a public good, therefore assumed to be ‘non-excludable’, ‘non-depletable’, and free to be used without limits by anyone, at any time, and anywhere across geographical boundaries. From such a perspective, it is apparent that knowledge and information are treated as being largely synonymous. However, once tacit knowledge is taken into consideration, it becomes clear that technology as a whole cannot be easily traded or exchanged, and only the potentially public knowledge component is liable to be assessed in terms of transaction cost analysis.

When narrowing the notion of technology to something akin to information, and concentrating on the organisation of the exchange of such information, there is a tendency to overemphasise the appropriability issue (Winter, 1987, 1993). The return on innovation to a firm may well be mainly a return on its creation of tacit capability, a process supported by, but not reducible to, the generation of new potentially public knowledge. In addition, knowledge can be at the same time both 'sticky' within the organisation or firm boundaries, while also being 'leaky' or mobile, generating outflows in the environment external to the firm (von Hippel, 1994; Wernerfelt, 1984). Ideas, inventions and practices that are unable to move within organisations, in some circumstances may prove to be quite capable of moving outside of them (Seely Brown and Duguid, 2001; Winter, 1987), thereby putting in question the centrality of the appropriability issue.

The main reason for knowledge to be confined to certain geographical contexts is assumed to be its inherent complexity – particularly with regard to technical knowledge – that may make it difficult to share among different interacting actors or organisations. Such complexity may prevent knowledge from being codified and made explicit and mobile, and thereby stored and transmitted by way of information. These arguments underlie the knowledge 'filter' hypotheses (Acs, 2002), and the wider notion of knowledge may also embrace cultural and institutional differences, which shape the spatial patterns of its functional dimensions, that is, knowledge production, absorption and diffusion.

By viewing knowledge-creation processes as complex, systemic, cumulative, partially tacit and sticky (whether codified or not) phenomena, there are strong grounds for arguing that innovation is very likely to stay highly concentrated across space, organisations and hierarchies, thereby giving rise to rather distinctive growth patterns. Technological innovation is in fact generally more 'sticky' than production, and this may be explained through the distinction – introduced by Henderson and Clark (1990) and developed within spatial systems by Phene and Tallman (2002) – between *component* and *architectural* knowledge. The first encompasses specific resources, assets, skills and technical systems that refer to particular constituents of the organisational system rather than to the whole. Architectural knowledge is instead related to an organisational system as a whole, which, through its institutional structure and routines, arranges the components for productive use (Phene and Tallman, 2002). Such architectural knowledge is cluster-specific, has a path-dependent and evolutionary nature and is developed through relationships at both inter-firm and intra-community level. Thus, it provides a common framework through which the creation, adaptation, and diffusion of component knowledge is highly facilitated, raising intra-cluster absorptive capacity and extra-cluster linkages – therefore attractiveness to MNEs – vis-à-vis that of other locations.

Regarding the third limit of the transaction cost view, we must also consider how each of the issues discussed here is related to the base and the sources of innovation in the industry. On this point, works based on Pavitt's seminal taxonomy (1984) – used to explain firm and sectoral patterns of technical change and innovation among innovative manufacturing sectors – have adapted the original framework to take into account structural changes, such as the emergence of information-intensive and life science-based firms, the increasingly blurred boundary between manufacturing and services, or the shifting of sectors among different patterns (see among others, Archibugi, 2001; Evangelista, 2000; Malerba and Orsenigo, 1996; Marsili, 2001; Pavitt et al., 1989).

More recent research, focusing on the determinants of cross-sectoral differences in agglomeration forces and dynamics, has emphasised the role of technological and learning processes that are 'likely to affect the relative importance of phenomena such as localised knowledge spillovers; inter- vs. intra-organizational learning; knowledge complementarities fuelled by localised labour mobility; innovative explorations undertaken through spin-offs, and more generally, the birth of new firms' (Bottazzi and Dindo, Chapter 24 in this book). Therefore, if we are to better relate questions of innovation to the emergence and evolution of clusters and regions, it is necessary to map knowledge and technology characteristics onto the simple transactions costs models described in Table 8.1.

The reason for this is primarily that the changes in the nature of knowledge and the emergence of new technical capabilities have been responsible for the transformation of the nature of the MNE from technology transferor to technology creator, and will determine how the logic of an MNE, as well as of a type of spatial agglomeration, will evolve over time.

5. An extended technology-based classification of spatial agglomeration variety

As is highlighted in the previous section, once we allow for a broader definition of knowledge, we can reasonably assume that the nature of agglomeration effects is likely to be highly sensitive not only to the industry structure, but also to the stage of product and cluster life cycle, and to changes in the underlying technological base (Audretsch, 1998; Boschma and Frenken, 2003; Breschi and Malerba, 2001). Several studies have emphasised the importance of specialisation in mastering a common knowledge set at the early stages of industry and cluster life cycles. Alternatively, diversity in complementary sets of competencies and knowledge has been argued to be more advantageous in later stages of growth of firms and local contexts, when interdependent pieces of knowledge have to be integrated and recombined to sustain innovation rates (see among others, Arora and Gambardella, 1994; Feldman, 1999; Frenken et al., 2007).

The evolution of firms in a specific industry and in a specific cluster is mainly shaped by the underlying knowledge conditions, the so-called 'technological regime', that is, a particular combination of appropriability (of the returns of innovation) conditions, technological opportunities (likelihood to innovate, given investments in research), degree of cumulativeness of technological knowledge (extent to which the amount of innovation produced in the past raises the probability of current innovation) and characteristics of the knowledge base (type of knowledge on which firm's activities are based). Technological regimes identify common properties of innovative processes in different sets of production activities, thus helping to explain asymmetries in the dynamics of industrial competition at sectoral and geographical level (among others Bottazzi et al., 2002; Dosi, 1988; Malerba and Orsenigo, 1995, 1996; Marsili, 2001; Nelson and Winter, 1982, and Bottazzi and Dindo, Chapter 24 in this book).

The shifts of technological regimes are also likely to have a geographical dimension, thereby affecting the distribution of innovative and technological activities across space (Breschi, 2000). The balance between internal and external sources of knowledge varies over time across and within innovating actors, either firms or other organisations. In fast-changing fields innovation processes are critically informed by new developments occurring outside any individual firm. Particularly in such cases, therefore, technological progress benefits greatly from the active participation in the technological community or context where new developments take place (Frost, 2001). Current technological opportunities affect the overall rate of regional technological growth, insofar as innovation flows will privilege, among core locations, those offering the best and fastest growing breaks. Indeed, the same concept of technological paradigm implies a close link between technical progress, organisation and socio-economic institutions: by definition, any radical innovation brings about to some extent transformations in the organisation of markets, production and communities. Therefore, organisational and institutional changes are inextricably associated with technological change, and diversity, both institutional and technical, characterises evolutionary development of socio-economic systems (Foray and Freeman, 1993).

As is discussed also in section 4, according to such an evolutionary perspective (which provides the basic rationale for revising the cluster classification in a dynamic sense), spatial typologies are not only heterogeneous, but also path dependent (Boschma and Lambooy, 1999; Cooke et al., 1997; Dopfer et al., 2004; Iammarino, 2005; Martin and Sunley, 2006, and Chapter 3 in this book). Historical contingency explains the actual existence of selection mechanisms in socio-economic processes: change is neither solely based on structural elements subject to general rules, nor purely driven by random effects. At each point in time in a system's evolution, a number of paths are theoretically possible, but only a few will be chosen by the actors because each path must conform to the logic of socio-economic dynamics (Schwerin and Werker, 2003; see also Hassink, Chapter 21 in this book).

Spatial configurations can thus be viewed as acting as selection mechanisms that may provide conditions favourable to meet the new requirements of technical change, that is, social capabilities for institutional change. Moreover, not only do the characteristics of the selection environment and their changes influence the actors, but the latter also change the environment (Cohen and Levinthal, 1990; Lambooy and Boschma, 2001). In this evolutionary view, growth opportunities are therefore assumed to be shaped and constrained by path dependency, that is, by the inheritance of local structural characteristics from past knowledge accumulation and learning, and these may often be geographically determined (Boschma, 2005; Boschma and Frenken, 2006; Malmberg, 1997).

Hence, to discuss the evolutionary dynamics of spatial agglomeration variety, the classification reported in Table 8.1 was extended in Table 8.2, in which the underlying knowledge conditions of the cluster are now made explicit. This allows us to take into account the ways in which firms, and particularly MNEs, may interact with the industrial and technological environments, and the multiple linkages between knowledge conditions and regional economic growth. Thereby, it may help understand how spillovers occur and change over time. Again, our classificatory attempt is process-driven, thereby it assumes that each typology may be dominant but not at all exclusive, allowing for geographical configurations that will actually show features of more than one category at the same time (even more so when considered in dynamic terms).

Table 8.2 Industrial clusters: knowledge, technology and cluster dynamics

Characteristics	Pure agglomeration	Industrial complex	Social network (SN)	
			New SN	Old SN
Nature of technical knowledge	Codified, explicit and mobile Transmitted by way of information	Mixed, systemic, routinised, R&D-intensive Specific, based on non-transferable experience	Tacit, new, generic, non-systemic, sticky and leaky Transmitted within cognitive networks	Mixed, mature, incremental Transmitted within localised networks
Technological trajectory	Oriented to processes, problem-solving	Oriented to complex products, cost-cutting	Oriented to radically new products	Oriented to processes, customer-driven
Dynamics	Stochastic	Strategic	Mixed	Mixed
Sources of innovation	External to the firm	Internal to the firm	Mixed	External to the firm
Appropriability of innovation returns	Low, perfect or monopolistic competition	High, private creation of new knowledge, oligopolistic competition	Mixed, public–private creation of new knowledge	Low, collaboration and competition
Technological opportunities	Variable	Low	Very high, uncertain	Low
Degree of cumulativeness	Low	High	Low	High
Knowledge-base	Diversified	Specialised	Research-based	Specialised along the filière
Modes of governance	Market	Hierarchies	Relational and cognitive networks	Social and historical networks
Examples of industrial specialisation	Finance, banking, insurance, business services, retailing	Steel, chemicals, automotive, pharmaceuticals, machine tools, medical instruments, ICT hardware	SME high-tech clusters in general purpose technologies	Customised traditional goods textiles, footwear, furniture, tourism
Example of cluster	'Silicon Valley' (California)	'Silicon Glen' (Scottish Electronics Industry)	'Silicon Fen' (Cambridge, UK)	Italian industrial districts (Emilia-Romagna)
Pavitt classification	Information-intensive, supplier dominated firms	Production-intensive firms (scale-intensive and specialised suppliers)	Science-based firms	Supplier-dominated firms

In the model of pure agglomeration, the bulk of knowledge is explicit and codified, available to any actor and organisation, and generated outside firms' boundaries, being largely created in public institutions. Variety and promiscuity are distinctive features of cities: the combination of different streams of knowledge occurs across a broad range of sectors (Jacobs externalities) and individual linkages or relations are unpredictable, because of the low degree of cumulativeness. However, even though in many cases the critical distance over which urban agglomeration externalities operate may be that of the broad city-metropolitan area (Gordon and McCann, 2005a, 2005b), as is assumed by many theoretical models of agglomeration, there is also much evidence to suggest that for many firm-types and industries, the critical distances over which agglomeration externalities operate may be very much larger than that of the city-region (Arita and McCann, 2000; Caniëls, 2000; Cantwell and Iammarino, 2003; Simmie, 1998; Suarez-Villa and Walrod, 1997). These considerations come from observations of regional innovation systems, the location and performance of MNE R&D facilities, the behaviour of local and regional labour markets, and the behaviour of transportation systems, particularly air-transport systems.

The combination of diverse kinds of knowledge into an interdependent economic and technological base needs crucially a plurality of sources and networking among them. The features of economic systems – and particularly their communication opportunities – play a major role in assessing the conditions of the production of new technology (Antonelli, 2000; Patrucco, 2001; see also Antonelli, Chapter 7 in this book). In this respect, urbanised and metropolitan regions have proved to offer highly positive institutional contexts explaining the features of the collective dynamic of technological progress, as a result of the mix of variety and complementarity of economic activities, science and technology infrastructures and communication and network mechanisms.[3] However, these findings are further complicated by the fact that while there is much evidence to suggest that the link between innovation and cities can be strong in certain sectors (Acs, 2002), the evidence on this particular issue is also not always conclusive. Cities do not always appear to be centres of innovation, and nor does innovation necessarily appear to be centred on cities (Simmie and Sennet, 1999).

As described in Table 8.2, the industrial complex model is instead associated primarily with cumulative learning from sources inside the industry and the firm, such as in-house R&D, and on the basis of knowledge that is specific to industrial applications. Such cases generally exhibit low entry possibilities and high industry concentration, which is likely to display a complementary strong concentration at the spatial level. Large incumbents, often MNEs, account for most of the sector's innovative activity, and these firms can profit from their innovations in part because they have the potential to exclude rival firms from using the new products and processes they have generated. In these situations, knowledge based on non-transferable experience is an important input in generating innovative activity, and the incumbent firms tend to have the innovative advantage over new firms because innovation is relatively routinised and processed within the existing hierarchical bureaucracies. As such, leading firms, particularly large MNEs, play a crucial role, and power asymmetry is central to the value chain and the governance of innovation in the cluster or region (Cooke, 2001). Regions with a strong specialised industrial structure and composite and advanced knowledge-production basis may provide the most suitable environment for technology creation and experimentation, as well as for the development of multilateral networks of dissimilar but complementary relations between MNEs and local actors (Patrucco, 2001, 2003).

On the other hand, from the perspective of inward investment, in the industries in which the host region or cluster is technologically strong, the vibrant presence of indigenous companies tends to deter MNEs in the same industry from conducting substantial levels of innovative activity in the primary technology of the industry in question. At the same time, the strongest firms of other industries, and particularly foreign MNEs, might be attracted to that location for those relevant technologies, as these lines of development for them represent diversification from the primary technologies of their own industry. Since they are in another industry, they are not direct competitors of the local leaders. Comparative strength in an industry of locally owned firms encourages outward investment in technological activity, but discourage the inward investment of foreign-owned companies in the same industry: MNEs' penetration in locations of domestic technological excellence tends to be low in most countries (Cantwell and Kosmopoulou, 2002)

Once the different types of spatial configuration are distinguished in terms of technological regimes, structures and governances, as well as their transactional relations, it becomes clear that the social network model ceases to be a relatively homogeneous and consistent analytical category. Thus, in Table 8.2 the single transaction cost-based social network typology is split into two sub-categories, namely the 'new social network' category and the 'old social network' category.

In the new social network model technological opportunities come in the main from sources outside the firm and the industry sector, such as academic research. In this kind of technological environment the type of knowledge tends to be both generic and non-systemic, there is a high rate of market entry and exit, a strong volatility of market shares, and a low level of market concentration. In this environment, the tacit and sticky nature of knowledge requires geographical proximity. On the other hand, the relative 'leaky' character of new knowledge – and the high potential for spillovers – the openness of the innovation system and the related emergence of new rules, standards, blueprints and verification procedures, all point to the importance of external sources of technical knowledge, which may not be necessarily localised. As such, innovation is frequently associated with a high level of uncertainty regarding both technology and demand, and a high level of market turbulence. As a consequence, a lower survival of new firms is likely to be associated with higher innovativeness and growth of surviving firms (Alchian, 1957). Innovations also therefore mainly come from knowledge that does not have a routinised nature, with new firm start-ups playing an important role, SMEs accounting for the bulk of innovative activity, and normally limited presence of large MNEs *in situ*, at least in the first stages of the cluster life cycle.

In the old social network type, there is not necessarily any clear hierarchical structure, and the overall coordination of the innovation system is left to a mix of cooperation and competition. Knowledge is largely codified and mature, it develops along trajectories that are mainly oriented to process innovation, and it is transmitted essentially by way of personal contacts, social and political lobbying, and backward and forward linkages, the latter sometimes intensified by the presence of MNEs.

Moreover, from the perspective of geography, there is also a fundamental difference in the particular systems of innovation governance between new and old social networks

(Simmie et al., 2004; von Tunzelmann, 2003). While old social networks are generally geographically embedded, and also rooted in historical experience implying strong social proximity, new social networks may rely on various kinds of communities of practice that do not necessarily require a spatial dimension, with a strong role played by relational and cognitive proximity (Boschma, 2005).

Some empirical analyses have shown that inter-industry knowledge spillovers are likely to become more intense in centres of technological excellence, where spillovers seem to operate mainly through exchanges in and around core technological systems (i.e. rooted in 'general purpose technologies' such as, for instance, background engineering, mechanical methods, electronics and ICTs), creating linkages between actors in quite separate alternative fields of specialisation. These centres of excellence – which are more likely to be classified either as pure agglomerations or as new social networks, and show different scope for MNE presence and contribution to spillovers – experience a faster process of convergence between old and new technologies and a potentially greater competitiveness, eventually leading to processes of rise and decline of technological clusters (Cantwell and Iammarino, 2003; see also Lambooy, Chapter 22 in this book).

6. The dynamic relationship between MNEs and innovative clusters

The arguments developed in the previous sections allow us to ask questions concerning how an observed industrial agglomeration may evolve over time. Once again, what is important here is to identify which of these idealised technological types best approximates the dominant characteristics of the geographical unit under analysis. In reality, it is likely that various mixed, diversified, and idiosyncratic patterns of growth can be observed, and no linear or deterministic development path can necessarily be established. Also, firms are nor necessarily clustered together in space; and clusters, where they do exist, are not necessarily innovative systems. However, in observed situations where agglomerations have emerged over time, some particular types of evolutionary or transition pattern are likely to be more common than others.

As widely highlighted in the literature on multinationals, the accumulation of technological competence is a path-dependent process, being partly firm-specific and partly location-specific. MNEs spread the competence base of the firm, and acquire new technological assets or sources of competitive advantage. The strategic internationalisation of technological operations has indicated that decisions on what and where to internationalise are strictly related to the roots of the firm's competitiveness. An effective approach to the strategic management of technological functions entails the evaluation of the core technological competence, that is, the set of knowledge, skills and capabilities that makes the MNE's innovative capacity unique and original: the locational choice is part of the strategy and a central issue to optimise technological effectiveness and growth (Chiesa, 1995).

The change in MNE strategies towards a growing degree of cross-location coordination of their operations requires an organisational structure that could not have been developed through purely arms-length market-based coordination between geographically separated units. Transaction costs may shrink with integrated governance of units, whatever units are considered to be, either affiliates or different locations (Dunning and Robson, 1987). Hence, the transaction cost approach, focusing basically on benefits in terms of short-term efficiency and flexibility, fails to take into account the knowledge accumulation effects linked to the integration of innovative activities across geographically dispersed units.

MNE technological experience is a fundamental ingredient of a cumulative process of development, as it has a positive and significant impact on the creation of knowledge linkages and of extensive networks of relationships with local firms endowed with complementary assets (Castellani and Zanfei, 2002). The presence of a number of leading foreign-owned companies tends to attract further knowledge, to stimulate spin-offs and to generate a positive cycle: once a cluster establishes itself as a technology hotspot, it can experience rapid and continuous growth. Localised knowledge stocks are thus very important, in so far as the cumulative nature of innovation will tend to make advantaged regions more advantaged compared to others in the next round of innovations on the basis of the accumulated knowledge stock (Beaudry and Breschi, 2000; Malerba, 1992).

Since technology is localised and context-specific, the technological trajectories of large MNEs tend to lock in to particular spatial configurations. However, the cumulative and incremental nature of technological change implies that the day-to-day adaptation of technology, through an interaction between innovative and production activities of firms, has a more pervasive effect than the major technological shifts, which generate entirely new products and production processes. In other words, to secure the long-term objectives of strategic competitiveness, MNEs need a current commitment to the regeneration of core technologies that can provide the basis for possible more radical innovation (Pearce and Papanastassiou, 2007).

The classification presented in Table 8.2 may offer a grid for interpreting the critical issue of the evolving 'global versus local' interaction. As is said above, all three categories are predicated on the existence of knowledge spillovers, which however differ in scope and nature across the spatial typologies, and are subject to shifts over time. A few brief examples can be used to illustrate the different directions of shifts in such interactions, and their impact on geographical agglomeration.

The early development of the global automobile industry on both sides of the Atlantic approximated to something akin to 'pure agglomeration' (Boschma and Wenting, 2007; Hall, 1998), but over time this system evolved to represent the classic 'industrial complex' model we now see (Best, 1990), which is dominated by large oligopolistic producers, clustered in particular localities, with complex and highly organised input–output supply chain systems (Markusen, 1996). The evolutionary process of this location-specific sector has therefore been primarily from pure agglomeration to industrial complex.

If we consider the transformation of the Italian industrial districts in the light of the process of internationalisation, we may indeed argue that 'the industrial district has often proved to be rather a "stage" in one of the possible different paths of industrialization' (Becattini, 1987, p. 32). As such, this case now provides examples of evolutionary transitions from 'old social networks' to something that exhibits many of the relational characteristics of the 'industrial complex' model, except for a greatly reduced geographic localisation of many of the input–output linkages often transferred into MNEs' global production networks (see for example Belussi and Sammarra, 2010; Garofoli, 2003). On the other hand, some recent empirical evidence seems to indicate the possibility for mature industrial regions characterised by spatial agglomerations of the 'old social network' type to undergo a process of revitalisation of technological capabilities towards new and research-intensive productions, shifting to models closer to 'new social networks' (Belussi et al., 2007).

There are also examples of rising industrial clusters in some newly emerging and rapidly changing industrial sectors such as biotechnology and multimedia (e.g. Baptista and Swann, 1998; Fuchs and Koch, 2005; Swann and Prevezer, 1996). Many of the innovations within these sectors take place in large MNEs whose locational criteria primarily reflect those of the 'industrial complex'. However, in situations where activities in these sectors are geographically concentrated among small and medium firms, they appear to correspond most closely to the 'new social network' type of system, where inter-industry spillovers emerge from the integration of different types of network, and where the local embeddedness of MNE competence-creating subsidiaries varies greatly among locations (Cantwell and Piscitello, 2005). However, the recent emergence of these geographically clustered industries means that as yet it is too early to point to a particular evolutionary path.

At this stage, however, it must be made clear that differences in cluster types and also in cluster evolutionary paths, where they exist, are not necessarily related to industrial sectors. In particular, the technology content dichotomy mentioned in section 2 clearly turns out to be mostly insignificant: high-skill, high-technology sectors do not exhibit particular cluster characteristics or evolutionary trajectories. In the case of the global semiconductor and electronics industry, for example, much of the industry initially emerged from oligopolistic firms in other sectors, such as defence contracting, lighting engineering, or radio- and telecommunications (Hall, 1998). The majority of the global semiconductor industry, involving wafer fabrication and assembly activities, is still largely dominated by large MNEs both in the USA, and in Europe and Asia. The location behaviour of these firms generally reflects rather traditional location criteria involving a consideration of location-specific factor costs and the transaction costs involved in coordinating business activities over space (McCann, 1995). As such, in situations where we observe firms from these industries to be clustered together in space, their location-organisation characteristics reflect primarily those of the 'industrial complex' model (Arita and McCann, 2002a, 2002b; McCann et al., 2002). These location-specific sectors emerged initially as an industrial complex, and have remained so for over fifty years. As such, no real cluster-evolutionary path is discernible in this case.

On the other hand, there are some sub-components of the global semiconductor and electronics industry that have emerged in quite different ways, for example, the Silicon Valley elements of this industry, which have tended to focus on semiconductor design activities. Although the early post-war features of the Silicon Valley semiconductor industry were mainly typical of the 'old social network' model (Hall, 1998; Saxenian, 1994), this industry initially developed during the 1970s along the lines of a 'new social network', and has now emerged into something that is akin to a 'pure agglomeration' model (Arita and McCann, 2000, 2004), exhibiting the supplier-dominated characteristics of Pavitt's classification. The majority of the design innovations developed in Silicon Valley are made possible essentially because of the miniaturisation innovations generated in the wafer processing and wafer assembly parts of the industry that are primarily located elsewhere, as components of global value chains with large MNE flagships. As such, the evolutionary transition of the Silicon Valley cluster has been from old social network to new social network to pure agglomeration.

Finally, in the case of the Scottish electronics industry, which is often known as 'Silicon Glen', and which specialises in the production of ICT equipment with a considerable incidence of foreign MNE research carried out *in situ*, this location-specific sub-component of the semiconductor and electronics industry emerged as an 'industrial complex' in the 1960s and 1970s, and has remained so for over forty years. As such, even taking into account the changes brought about by technical advances in the industry over time, no real evolutionary path is discernible in this particular cluster in terms of spatial configuration. Meanwhile, if we consider the case of the high technology cluster of electronics firms around Cambridge, UK (Castells and Hall, 1994), the emergence of this location-specific sector can be characterised by a movement from an 'old social network' to a 'new social network'. The system is still far too small to be really considered an agglomeration along the lines of Silicon Valley.

7. Conclusions

The possible alternative characteristics of clusters presented here indicate that neither technological or knowledge features alone are a sufficient guide to the types of spatial configuration that are likely to emerge, nor are the nature of transactions or industry characteristics. Rather, as we have seen, knowledge and innovation processes, organisational modes, firm- and industry-specific characteristics, institutional and governance settings, all play a role in explaining the diversity of industrial agglomerations and also their evolutionary trajectories. Indeed, as any single firm, particularly when large and multinational, can follow more than one technological trajectory (Pavitt et al., 1989), clusters may well be engaged in a prevalent but not exclusive trajectory at any given point of time. Process-based classificatory attempts, such as that presented in this chapter, help thus to explain multiple trajectories and patterns of evolution.

When considering the disruptive changes imposed by the growing interdependence between the 'global' and the 'local', where localised production systems may correspond to globalised knowledge systems, the potential dynamic advantages of clustering for MNEs (and vice versa) can only be grasped by adopting evolutionary views on technological change and regional growth. Path-dependent processes shape cluster features and their variety, as they do with firm heterogeneity. In exploring geographical variations in local absorptive capacity, and the potential for MNEs to tap into the local knowledge base fostering spillovers, it is necessary to look beyond firms' competence accumulation and to consider the likelihood of endogenous or evolutionary 'feedback' mechanisms actually operating in local institutional settings (von Tunzelmann, 2009).

Evolutionary approaches to clustering and MNEs still face major challenges. Here, we would like to mention two in particular, which coincide with our next research steps. The first one is that, once we account for path-dependent innovation and knowledge creation processes, it becomes very difficult to apply simple stylised cluster constructs, because there is neither a representative Marshallian firm nor a typical 'innovative' cluster. In terms of policy analysis, therefore, applying ex ante categories to clustering observations is not a sufficiently analytical standpoint. More work is surely needed in the direction of providing the basis for policy guidelines aimed at improving the likelihood of local innovation and its attractiveness for MNEs in a variety of industries, locations and countries.

Second, and relatedly, our suggestions on cluster dynamics clearly need further exploration. The features of spatial concentration development, the probability of some evolutionary patterns relative to others, the prospects for such patterns (or for the emergence of new ones) in the light of speeding globalisation and technical change, all require both additional theoretical consideration and sound empirical research, in the effort to avoid the appealing temptation of simplistic determinism.

Notes

1. It is beyond the scope of this chapter to engage with the ongoing debate on the relevant spatial scale for economic analysis and, as a consequence, with the definition of 'cluster'. See, for example, Maskell (2001), Martin and Sunley (2003) and Moulaert and Sekia (2003).
2. For a detailed discussion of each category see Gordon and McCann (2000).
3. These considerations can help explain the extreme spatial concentration of MNE headquarters in the major world cities, where transactions and networks are strongly information- and power-intensive, and institutional connectivity is dense enough to make corporate networks for production and technology work properly elsewhere (see for example Friedmann, 1986; Sassen, 2001).

References

Acs, Z. (2002), *Innovation and the Growth of Cities*, Cheltenham, UK and Northampton, MA, USA: Edward Elgar.

Akerlof, G.A. (1970), 'The market for lemons: quality uncertainty and the market mechanism', *Quarterly Journal of Economics*, **84**, 488–500.

Alchian, A.A. (1957), 'Uncertainty, evolution and economic theory', *Journal of Political Economy*, **58**, 211–21.

Almeida, P. (1997), 'Knowledge sourcing by foreign multinationals: patent citation analysis in the US semiconductor industry', *Strategic Management Journal*, **17**, 155–65.

Antonelli, C. (2000), 'Collective knowledge communication and innovation: the evidence of technological districts', *Regional Studies*, **34**, 535–47.

Archibugi, D. (2001), 'Pavitt's taxonomy sixteen years on: a review article', *Economics of Innovation and New Technology*, **10**, 415–25.

Arita, T. and P. McCann (2000), 'Industrial alliances and firm location behaviour: some evidence from the US semiconductor industry', *Applied Economics*, **32**, 1391–403.

Arita, T. and P. McCann (2002a), 'The spatial and hierarchical organization of Japanese and US multinational semiconductor firms', *Journal of International Management*, **8** (1), 121–39.

Arita, T. and P. McCann (2002b), 'The relationship between the spatial and hierarchical organization of multiplant firms; observations from the global semiconductor industry', in P. McCann (ed.), *Industrial Location Economics*, Cheltenham, UK and Northampton, MA, USA: Edward Elgar.

Arita, T. and P. McCann (2004), 'Industrial clusters and regional development: a transactions costs perspective on the semiconductor industry', in H.L.F. de Groot, P. Nijkamp and R.R. Stough (eds), *Entrepreneurship and Regional Economic Development: A Spatial Perspective*, Cheltenham, UK and Northampton, MA, USA: Edward Elgar.

Arora, A. and A. Gambardella (1994), 'The changing technology of technological change: general and abstract knowledge and the division of innovative labour', *Research Policy*, **23**, 523–32.

Audretsch, D.B. (1997), 'Technological regimes, industrial demography and the evolution of industrial structures', *Industrial and Corporate Change*, **6** (1), 49–82.

Audretsch, D.B. (1998), 'Agglomeration and the location of innovative activity', *Oxford Review of Economic Policy*, **14** (2), 18–29.

Autant-Bernard, C., V. Mangematin and N. Massard (2003), 'Creation and growth of high-tech SMEs: the role of the local environment', in J. Leage-Hellman, M. McKelvey and A. Rickne (eds), *The Economic Dynamics of Biotechnology: Europe and Global Trends*, Cheltenham, UK and Northampton, MA, USA: Edward Elgar.

Baptista, R. and P. Swann (1998), 'Do firms in clusters innovate more?', *Research Policy*, **27** (5), 525–40.

Beaudry, C. and S. Breschi (2000), 'Does "clustering" really help firms' innovative activities?', Manchester Business School, Working Paper 412.

Becattini, G. (ed.) (1987), *Mercato e forze locali. Il distretto industriale*, [Market and Local Forces. The Industrial District], Bologna: Il Mulino.

Bell, M. and M. Albu (1999), 'Knowledge systems and technological dynamism in industrial clusters in developing countries', *World Development*, **27** (9), 1715–34.

Belussi, F. and A. Sammarra (2010), *Business Networks in Clusters and Industrial Districts. The Governance of the Global Value Chain*, London & New York: Routledge.

Belussi, F., A. Sammarra and S.R. Sedita (2007), 'Learning at the boundaries in an "open regional innovation system": a focus on firms' innovation strategies in the Emilia Romagna life science industry', mimeo.

Best, M.H. (1990), *The New Competition: Institutions of Industrial Restructuring*, Cambridge: Polity Press.

Boschma, R.A. (2005), 'Proximity and innovation: a critical assessment', *Regional Studies*, **39** (1), 61–74.

Boschma, R.A. and K. Frenken (2003), 'Evolutionary economics and industry location', *Review for Regional Research (Jahrbuch fur Regionalwissenschaft)*, **23**, 183–200.

Boschma, R.A. and K. Frenken (2006), 'Why is economic geography not an evolutionary science? Towards an evolutionary economic geography', *Journal of Economic Geography*, **6**, 273–302.

Boschma, R.A. and J.G. Lambooy (1999), 'Evolutionary economics and economic geography', *Journal of Evolutionary Economics*, **9**, 411–29.

Boschma, R.A. and R. Wenting (2007), 'The spatial evolution of the British automobile industry. Does location matter?', *Industry and Corporate Change*, **16** (2), 213–38.

Bottazzi, G., G. Dosi and G. Fagiolo (2002), 'Mapping sectoral of technological accumulation into the geography of corporate locations. a simple model and some promising evidence', LEM WP 2002/21, Sant'Anna School of Advanced Studies.

Breschi, S. (2000), 'The geography of innovations: a cross-sector analysis', *Regional Studies*, **34** (3), 213–29.

Breschi, S. and F. Lissoni (2001), 'Knowledge spillovers and local innovation systems: a critical survey', *Industrial and Corporate Change*, **10** (4), 975–1005.

Breschi, S. and F. Malerba (2001), 'The geography of innovation and economic clustering: some introductory notes', *Industrial and Corporate Change*, **10** (4), 817–34.

Caniëls, M.C.J. (2000), *Knowledge Spillovers and Economic Growth: Regional Growth Differentials Across Europe*, Cheltenham, UK and Northampton, MA, USA: Edward Elgar.

Cantwell, J.A. (1989), *Technological Innovation and Multinational Corporations*, Oxford: Basil Blackwell.

Cantwell, J.A. (ed.) (1994). *Transnational Corporations and Innovatory Activities*, London: Routledge.

Cantwell, J.A. (1995), 'The globalisation of technology: what remains of the product cycle model?', *Cambridge Journal of Economics*, **19** (1), 155–74.

Cantwell, J.A. and S. Iammarino (1998), 'MNCs, technological innovation and regional systems in the EU: some evidence in the Italian case', *International Journal of the Economics of Business*, Special Issue, **5** (3), 383–408.

Cantwell, J.A. and S. Iammarino (2003), *Multinational Corporations and European Regional Systems of Innovation*, London and New York: Routledge.

Cantwell, J.A. and O.E.M. Janne (1999), 'Technological globalisation and the innovative centres: the role of corporate technological leadership and locational hierarchy', *Research Policy*, **28** (2–3), 119–44.

Cantwell, J.A. and E. Kosmopoulou (2002), 'What determines the internationalisation of corporate technology?', in V. Havila, M. Forsgren and H. Håkanson (eds), *Critical Perspectives on Internationalisation*, Oxford: Pergamon, pp. 305–334.

Cantwell, J.A. and L. Piscitello (2005), 'Recent location of foreign-owned R&D activities by large multinational corporations in the European regions: the role of spillovers and externalities', *Regional Studies*, **39** (1), 1–16.

Cantwell, J.A. and G.D. Santangelo (1999), 'The frontier of international technology networks: sourcing abroad the most highly tacit capabilities', *Information Economics and Policy*, **11**, 101–23.

Castellani, D. and A. Zanfei (2002), 'Multinational experience and the creation of linkages with local firms: evidence from the electronic industry', *Cambridge Journal of Economics*, **26** (1), 1–25.

Castells, M. and P. Hall (1994), *Technopoles of the World: The Making of the 21st Century Industrial Structures*, London: Routledge.

Chandler, A.D., P. Hagström and O. Sölvell (eds) (1998), *The Dynamic Firm. The Role of Technology, Strategy, Organisation and Regions*, Oxford: Oxford University Press.

Chesnais, F. (1988), 'Technical co-operation agreements between firms', *Science Technology Industry Review*, **4**, 57–119.

Chiesa, V. (1995), 'Globalizing R&D around centres of excellence', *Long Range Planning*, **28** (6), 19–28.

Cohen, W.M. and D.A. Levinthal (1990), 'Absorptive capacity: a new perspective on learning and innovation', *Administrative Science Quarterly*, **35**, 128–52.

Cooke, P. (2001), 'Regional innovation systems, clusters, and the knowledge economy', *Industrial and Corporate Change*, **10** (4), 945–74.

Cooke, P., M. Gomez Uraga and G. Etxebarria (1997), 'Regional innovation systems: institutional and organisational dimensions', *Research Policy*, **26**, 475–91.

Dopfer, K., J. Foster and J. Potts (2004), 'Micro–meso–macro', *Journal of Evolutionary Economics*, **14**, 263–79.

Dosi, G. (1988), 'Sources, procedures and microeconomic effects of innovation', *Journal of Economic Literature*, **26**, 1120–71.

Dosi, G., F. Malerba O. Marsili and L. Orsenigo (1997), 'Industrial structures and dynamics: evidence, interpretations and puzzles', *Industrial and Corporate Change*, **6** (1), 3–24.
Dunning, J.H. (1970), *Studies in International Investment*, London: Allen & Unwin.
Dunning, J.H. (1993), *Multinational Enterprises and the Global Economy*, Wokingham: Addison-Wesley.
Dunning, J.H. and P. Robson (eds) (1987), *Multinationals and the European Community*, Oxford: Basil Blackwell.
Enright, M. (1998), 'Regional clusters and firm strategy', in A.D. Chandler, P. Hagström and O. Sölvell (eds), *The Dynamic Firm. The Role of Technology, Strategy, Organization and Regions*, Oxford: Oxford University Press.
Ernst, D., P. Guerrieri, S. Iammarino and C. Pietrobelli (2001), 'SME clusters facing global restructuring: what can be learnt comparing Italy and Taiwan?', in P. Guerrieri, S. Iammarino and C. Pietrobelli (eds), *The Global Challenge to Industrial Districts. The Case of Italy and Taiwan*, Cheltenham, UK and Northampton, MA, USA: Edward Elgar.
Evangelista, R. (2000), 'Sectoral patterns of technological change in services', *Economics of Innovation and New Technology*, **9** (3), 183–221.
Feldman, M.P. (1999), 'The new economics of innovation, spillovers and agglomeration: review of empirical studies', *Economics of Innovation and New Technology*, **8** (1), 5–25.
Foray, D. and C. Freeman (eds) (1993), *Technology and the Wealth of Nations*, London: Pinter.
Fors, G. (1998), 'Locating R&D abroad: the role of adaptation and knowledge-seeking', in P. Braunerhjelm and K. Ekholm (eds), *The Geography of Multinational Firms*, Boston, Dordrecht and London: Kluwer Academic Publishers.
Frenken, K., F.G. van Oort and T. Verburg (2007), 'Related variety, unrelated variety and regional economic growth', *Regional Studies*, **41** (5), 685–97.
Friedmann, J. (1986), 'The world city hypothesis', *Development and Change*, **17**, 69–83.
Frost, T.S. (2001), 'The geographic sources of foreign subsidiaries' innovations', *Strategic Management Journal*, **22**, 101–23.
Fuchs, G. and A. Koch (2005), 'Biotechnology and multimedia: cluster development in new industries', in B. Johansson, C. Karlsson and R.R. Stough (eds), *Entrepreneurship, Spatial Industrial Clusters and Inter-Firm Networks*, Cheltenham, UK and Northampton, MA, USA: Edward Elgar.
Fujita, M. P. Krugman and A.J. Venables (1999), *The Spatial Economy*, Cambridge, MA: MIT Press.
Garofoli, G. (2003), 'Distretti industriali e processo di globalizzazione: trasformazioni e nuove traiettorie', [Industrial districts and the globalization process: transformation and new trajectories], in G. Garofoli (ed.), *Impresa e territorio*, [Enterprise and Territory], Bologna: Il Mulino.
Giuliani, E. (2005), 'The structure of cluster knowledge networks: uneven and selective, not pervasive and collective', *DRUID Working Paper* 2005–11.
Glaeser, E.L., H.D. Kallal, J.A. Schinkmann and A. Shleifer (1992), 'Growth in cities', *Journal of Political Economy*, **100**, 1126–52.
Gordon, I.R. and P. McCann (2000), 'Industrial clusters: complexes, agglomeration and/or social networks?', *Urban Studies*, **3**, 513–32.
Gordon, I.R. and P. McCann (2005a), 'Innovation, agglomeration and regional development', *Journal of Economic Geography*, **5** (5), 523–45.
Gordon, I.R. and P. McCann (2005b), 'Clusters, innovation and regional development: an analysis of current theories and evidence', in B. Johansson, C. Karlsson and R.R. Stough (eds), *Entrepreneurship, Spatial Industrial Clusters and Inter-Firm Networks*, Cheltenham, UK and Northampton, MA, USA: Edward Elgar.
Granovetter, M. (1973), 'The strength of weak ties', *American Journal of Sociology*, **78**, 1360–80.
Granovetter, M. (1985), 'Economic action and social structure: the problem of embeddedness', *American Journal of Sociology*, **91**, 481–510.
Guerrieri, P. and C. Pietrobelli (2004), 'Industrial districts' evolution and technological regimes: Italy and Taiwan', *Technovation*, **24**, 899–914.
Guy, F. (2009), *The Global Environment of Business*, Oxford: Oxford University Press.
Hall, P. (1998), *Cities of Tomorrow*, Oxford: Blackwell.
Henderson, J.V., A. Kuncoro and M. Turner (1995), 'Industrial development in cities', *Journal of Political Economy*, **103**, 1067–85.
Henderson, R. and K. Clark (1990), 'Architectural innovation', *Administrative Science Quarterly*, **35**, 9–30.
Hilpert, U. (ed.) (2003), *Regionalisation of Globalised Innovation. Locations for Advanced Industrial Development and Disparities in Participation*, London: Routledge.
Iammarino, S. (2005), 'An evolutionary integrated view of regional systems of innovation. Concepts, measures and historical perspectives', *European Planning Studies*, **13** (4), 495–517.
Jacobs, J. (1960), *The Economy of Cities*, New York: Random House.
Jaffe, A.B., M. Trajtenberg and R. Henderson (1993), 'Geographic localization of knowledge spillovers as evidenced by patent citations', *Quarterly Journal of Economics*, **63** (3), 577–98.

Kogut, B. and S.J. Chang (1991), 'Technological capabilities and Japanese foreign direct investment in the United States', *Review of Economics and Statistics*, **73** (3), 401–13.
Krugman, P. (1991), 'Increasing returns and economic geography', *Journal of Political Economy*, **99** (33), 483–99.
Kuemmerle, W. (1999), 'The drivers of foreign direct investment into research and development: an empirical investigation', *Journal of International Business Studies*, **30** (1), 1–24.
Lambooy, J.G. and R.A. Boschma (2001), 'Evolutionary economics and regional policy', *The Annals of Regional Science*, **35**, 113–31.
McCann, P. (1995), 'Rethinking the economics of location and agglomeration', *Urban Studies*, **32** (3), 563–77.
McCann, P. (2001), *Urban and Regional Economics*, Oxford: Oxford University Press.
McCann, P., T. Arita and I.R. Gordon (2002), 'Industrial clusters, transactions costs and the institutional determinants of MNE behaviour', *International Business Review*, **11** (6), 647–63.
McCann, P. and R. Mudambi (2004), 'The location behaviour of the multinational enterprise: some analytical issues', *Growth and Change*, **35** (4), 491–524.
McCann, P. and R. Mudambi (2005), 'Analytical differences in the economics of geography: the case of the multinational firm', *Environment and Planning A*, **37**, 1857–76.
McCann, P. and D. Shefer (2004), 'Location, agglomeration and infrastructure', *Papers in Regional Science*, 83 (1), 177–96.
McCann, P. and S. Sheppard (2003), 'The rise, fall and rise again of industrial location theory', *Regional Studies*, **37** (6 & 7), 649–63.
Malerba, F. (1992), 'Learning by firms and incremental technical change', *Economic Journal*, **102**, 845–59.
Malerba, F. and L. Orsenigo (1995), 'Schumpeterian patterns of innovation', *Cambridge Journal of Economics*, **19**, 47–65.
Malerba, F. and L. Orsenigo (1996), 'The dynamics and evolution of firms', *Industrial and Corporate Change*, **5** (1), 51–88.
Malmberg, A. (1997), 'Industrial geography: location and learning', *Progress in Human Geography*, **21**, 573–82.
Markusen, A. (1996), 'Sticky places in slippery space: a typology of industrial districts', *Economic Geography*, **72**, 293–313.
Marshall, A. (1920), *Principles of Economics*, 8th edn, London: Macmillan.
Marsili, O. (2001), *The Anatomy and Evolution of Industries: Technological Change and Industrial Dynamics*, Cheltenham, UK and Northampton, MA, USA: Edward Elgar.
Martin, P. and G.I.P. Ottaviano (1999), 'Growing locations: industry location on a model of endogenous growth', *European Economic Review*, **43**, 281–302.
Martin, R. and P. Sunley (2006), 'Path dependence and regional economic evolution', *Journal of Economic Geography*, **6**, 395–437.
Martin, R. and P. Sunley (2003), 'Deconstructing clusters: chaotic concept or policy panacea?', *Journal of Economic Geography*, **3** (1), 5-35.
Maskell, P. (2001), 'Towards a knowledge-based theory of the geographical cluster', *Industrial and Corporate Change*, **10** (4), 921–44.
Moulaert, F. and F. Sekia (2003), 'Territorial innovation models: a critical survey', *Regional Studies*, **37** (3), 289–302.
Nelson, R.R. (1991), 'Why do firms differ and how does it matter?', *Strategic Management Journal*, **14**, 61–74.
Nelson, R.R. and S.G. Winter (1982), *An Evolutionary Theory of Economic Change*, Cambridge, MA: Harvard University Press.
Ohmae, K. (1985), *Triad Power: The Coming Shape of Global Competition*, New York: The Free Press.
Patrucco, P.P. (2001), 'The production and communication of technological knowledge at the regional level: the European evidence', in B. Stanford-Smith and E. Chiozza (eds), *E-work and E-commerce: Novel Solutions and Practices for a Global Networked Economy*, (Vol. 1), Amsterdam: IOS Press.
Patrucco, P.P. (2003), 'Institutional variety, networking and knowledge exchange: communication and innovation in the case of the Brianza technological district', *Regional Studies*, **37** (2), 159–72.
Pavitt, K. (1984), 'Sectoral patterns of technical change: towards a taxonomy and a theory', *Research Policy*, **13**, 343–73.
Pavitt, K., M. Robson and J. Townsend (1989), 'Technological accumulation, diversification and organisation in UK companies, 1945–83', *Management Science*, **35**, 81–99.
Pearce, R.D. (1999), 'Decentralised R&D and strategic competitiveness: globalised approaches to generation and use of technology in multinational enterprises', *Research Policy*, **28** (2), 157–78.
Pearce, R.D. and M. Papanastassiou (2007), 'Multinationals and national systems of innovation: strategy and policy issues', presented at SPRU, University of Sussex, 24 October.
Phene, A. and S. Tallman (2002), 'The effects of regional clusters on knowledge stocks and flows: evidence from the biotechnology industry', University of Utah, mimeo.

Porter, M.E. (1990), *The Competitive Advantage of Nations*, Basingstoke & London: Macmillan.
Rugman, A. and A. Verbeke (2002), 'Multinational enterprises and clusters: an organising framework', mimeo.
Rugman, A.M. (2005), *The Regional Multinationals*, Cambridge: Cambridge University Press.
Sassen, S. (2001), *The Global City: New York, London, Tokyo*, Princeton, NJ: Princeton University Press.
Saxenian, A. (1994), *Regional Advantage. Culture and Competition in Silicon Valley and Route 128*, Cambridge, MA & London: Harvard University Press.
Schwerin, J. and C. Werker (2003), 'Learning innovation policy based on historical experience', *Structural Change and Economic Dynamics*, **14**, 385–404.
Seely Brown, J. and P. Duguid (2001), 'Knowledge and organisation: a social-practice perspective', *Organisation Science*, **12** (2), 198–213.
Simmie, J. (1998), 'Reasons for the development of "islands of innovation": evidence from Hertfordshire', *Urban Studies*, **35** (8), 1261–89.
Simmie, J. and J. Sennet (1999), 'Innovative clusters: theoretical explanations and why size matters', *National Institute Economic Review*, **4** (170), 87–98.
Simmie, J., C. Siino, J.M. Zuliani, G. Jalabert and S. Strambach (2004), 'Local innovation system governance and performance: a comparative analysis of Oxfordshire, Stuttgart and Toulouse', *International Journal of Technology Management*, **28** (3–6), 534–59.
Suarez-Villa and W. Walrod (1997), 'Operational strategy, R&D, and intra-metropolitan clustering in a polycentric structure: the advanced electronics industries of the Los Angeles Basin', *Urban Studies*, **34** (9), 1343–80.
Swann, P. and M. Prevezer (1996), 'A comparison of the dynamics of industrial clustering in computing and biotechnology', *Research Policy*, **25**, 1139–57.
Vernon, R. (1966), 'International investment and international trade in the product cycle', *Quarterly Journal of Economics*, **80** (2), 190–207.
von Hippel, E. (1994), 'Sticky information and the locus of problem solving: implications for innovation', *Management Science*, **40**, 429–39
von Tunzelmann, N. (2003), 'Historical coevolution of governance and technology in the industrial revolutions', *Structural Change and Economic Dynamics*, **14**, 365–84.
von Tunzelmann, N. (2009), 'Regional capabilities and industrial regeneration', in P. McCann, M. Farshchi and O. Janne (eds), *Technological Change and Mature Industrial Regions: Firms, Knowledge and Policy*, Cheltenham UK and Northampton, MA, USA: Edward Elgar.
Wernerfelt, B. (1984), 'A resource-based view of the firm', *Strategic Management Journal*, **5**, 171–80.
Williamson, O.E. (1975), *Markets and Hierarchies*, New York: Free Press.
Winter, S.G. (1987), 'Knowledge and competence as strategic asset', in D.J. Teece (ed.), *The Competitive Challenge: Strategies for Industrial Renewal*, Cambridge, MA: Ballinger Publishing.
Winter, S.G. (1993), 'On Coase, competence and the corporation', in O.E. Williamson and S.G. Winter (eds), *The Nature of the Firm*, Oxford: Oxford University Press.
Zander, I. (1999), 'Whereto the multinational? Evolution of technological capabilities in the multinational network', *International Business Review*, **8**, 261–91.

9 Emergence of regional clusters: the role of spinoffs in the early growth process

Michael S. Dahl, Christian R. Østergaard and Bent Dalum

1. Introduction

The literature on regional clusters has increased considerably during the last decade. Most of the theories focus on explaining their growth, internal dynamics and structure. The emergence and growth patterns are usually explained by such factors as unique local culture, regional capabilities, tacit knowledge or the existence of location-specific externalities (knowledge spillovers, networks, labour market pooling and specialised suppliers). However, these factors are not sufficient to explain the early formation of clusters, when it grows from one to many firms. The location-specific externalities are non-existing when there is only one company, that is, there are no labour market pooling externalities or knowledge spillovers. Additional firms have to enter the cluster before networks between employees working in various cluster-firms are formed, before knowledge spillovers occur and before firms can benefit from labour market pooling.

The dominant theories focus more on explaining ex post dynamics of clusters than their early development. This chapter focuses on the early phase and uses an alternative approach to analyse the role of geography in the formation of clusters. Three key determinants are identified: (1) the geographical dimension of entrepreneurial activity, (2) spinoffs from successful firms and (3) new market opportunities. The chapter studies the evolution of the wireless communications cluster in Northern Denmark and compares it with the evolution of other clusters.

Section 2 presents the theoretical framework on the emergence of clusters with discussions of the geographical dimension of entrepreneurial activity, the role of spinoffs and market growth. Section 3 describes the methodology. The history of the wireless cluster is described in section 4. Section 5 gives an overview of the generation of new firms in the cluster. This evidence is discussed and compared with studies of four other clusters in section 6, followed by the conclusions in section 7.

2. The emergence of clusters

The formation of clusters consists of several phases and the role of geography is changing during the evolution and over the industry life cycle. The process of clustering is not deterministic and might stop at various stages. The first firm in a cluster can, if successful, grow. It can be unsuccessful and close. However, a single firm is not a cluster. By definition numerous firms are needed. Firms can be attracted to the region, local firms can diversify from other industries, or new firms can be founded. The entry of additional firms is a key process in the emergence of clusters, where few firms grow to many. This process of early period evolution from a single firm to a cluster has, however, only received little attention in the literature.

Alfred Marshall (1920) observed that firms within the same industry often cluster in particular localities. He was among the first to provide rather detailed descriptions of the sources of concentration of specialised industries in particular localities. Based on his work, the emergence of clusters has been explained by the existence of location-specific externalities, such as (1) economies of specialisation caused by a concentration of firms being able to attract and support specialised suppliers, (2) economies of labour pooling, where the existence of a labour force with particular knowledge and skills attracts firms, which in turn attract and create more specialised labour, and (3) knowledge externalities, where knowledge and information flow more easily between actors located in a cluster than over long distances, which all lead to the existence of increasing returns to scale for firms located in clusters. Other factors have been added, such as unique local culture, regional capabilities, tacit knowledge, and favourable factor conditions. Porter (1998) explains the emergence by interaction between parts of his diamond (context for firm strategy and rivalry, factor conditions, demand conditions, and related and supporting industries) that are present in a location because of historical circumstances. He adds that chance events seen in its historical and locational context are important for the birth of clusters. As a result, the early growth process remains exceptional for each cluster.

Other studies have searched for the explanation in the founding of the first firm, with a focus on why the founder decided to set up a firm in a particular location. Accordingly, each cluster becomes a unique story, which makes it harder to generalise from anecdotal stories of founders' personal preferences, luck and 'foresightedness'. These studies often take a long look back in history. An example is the analysis of the French plastics injection-moulding cluster in Oyannax that can be traced back to the year 630, when the city was given a monopoly to make wooden combs (van der Linde, 2003). In the nineteenth century, they shifted from wooden to celluloid combs. In the 1930s, combs fell out of fashion and some unemployed craftsmen started production of plastics toys, and so on. In 1936, they adopted injection-moulding and became a French centre of plastics manufacturing. However, this story tells us very little about how and why the plastics cluster emerged. The problem with tracing back to the initial event that started a process that led to clustering is that the explanation often becomes colourful, but the causality remains blurred. In the case of Oyannax, the link between the regional industrial history of comb-making and the plastics cluster could be an effect of the geographical dimension of entrepreneurial activity, since entrepreneurs tend to start up their business where they live (Sorenson, 2003).

The distinction between what are the driving forces behind the emergence of clusters and what are related to the evolution of regional capabilities is not clear. According to Maskell et al. (1998), each region has a set of capabilities that consist of the institutional background, the structure of the industry, natural resources, knowledge and skills. These have been developed through a historical interactive process and further evolution relies on the creation, utilisation and reproduction of knowledge. Firms in an emerging cluster influence the region through creation of and demand for skills and knowledge, but the firms are also affected by the existing structure and institutions in the region (Storper, 1997; Storper and Walker, 1989).

Following this argument, the founding of the first cluster firm is often seen as being set in a particular geographical location by chance (Arthur, 1990). Historical and chance events would have provided a location structure; and that inherited structure combined with agglomeration tendencies would determine the future settlements in a region. New industries will be laid down layer by layer on inherited structures through the phases of development (Arthur, 1994). In an evolutionary perspective, agglomeration can be interpreted as the mechanism by which existing organisations breed the new ones founded by entrepreneurs.

This argument is a quite similar to the cluster theories that focus on location-specific externalities to cause industries to cluster spatially. However, these ex post dynamics can only exist when there are several firms located in a particular geographical location. The externalities may or may not be important for clusters, but they can not drive the emergence of clusters in the early phase.

The open window of locational opportunity (OWLO) model (Boschma and Knaap, 1999) distinguishes between two stages in the spatial evolution of a new industry. In the initial stage the location-specific externalities do not exist, while they may lead to clustering in the second stage. The OWLO model argues that the novelty and discontinuous nature of new industrial activities would imply that local supply of knowledge is unlikely to determine where the new industry would evolve. Some regions may provide some generic conditions that create a stimulating environment for new industries while others are non-stimulating, but the new industry may initially emerge in both types of region (Boschma and Knaap, 1999). As a result, chance events become important for the spatial evolution. The new industry also relies on the creative ability to transform generic conditions into specific conditions. Thus, the new industry uses its creative ability to create the required labour and technological knowledge or to attract it from other regions (Boschma and Knaap, 1999; Storper and Walker, 1989). However, creation of the necessary labour through internal training and learning effects does not take place without firms. Likewise, skilled labour is unlikely to be attracted to a region with very few firms. The evolution from one to many firms is by definition a key process in the emergence of clusters.

Geography has a different role than proposed by much of the existing cluster literature that focuses on the location-specific externalities. Geography plays a role in the location decision of entrepreneurs because they may tend to found their new firms within close proximity to their previous employers or close to where they reside and have social contacts and family, and so on. It requires a considerable effort to obtain the information and resources needed to base a firm in another location (Sorenson, 2003).

This does not necessarily imply that firms always will be located within close proximity to their past employer and residence. There are many examples of founders who search for the most suitable location among regions that either provide access to a large local market or, perhaps more important, offer the best selection of resources to the organisation. It is hard to argue that potential founders only have knowledge about their own local environment and the local entrepreneurial opportunities (Romanelli and Schoonhoven, 2001). Nonetheless, they argue that most new firms will be founded in the same region or very close to the firm that produced the entrepreneur, since social and economic forces will induce entrepreneurs to locate close to their origins, so they can maintain their social ties and continue exploiting their localised knowledge of venture capitalists, potential employees, and suppliers.

It requires a wide selection of resources and social commitments to found a firm (Stinchcombe, 1965). These resources, such as available technical personnel and financial resources, generally tend to be immobile and unevenly distributed across geographical space. Consequently, entrepreneurs are most likely to be tied to the region where they have useful social relations, even if another region is otherwise more attractive (Sorenson and Stuart, 2001). In general, we may expect that clusters of new firms in a particular industry continuously evolve in regions where human and other resources are abundantly present and where entrepreneurs are 'produced' on a large scale in the incumbent firms (Sorenson and Audia, 2000). Geographical immobility of entrepreneurs would then make the existing structural base of a region a dominant source of clusters and create an agglomeration effect similar to the ones discussed by Arthur (1990, 1994).

Jane Jacobs (1969) described how clusters often are driven by employees leaving incumbents to start new firms. She described how an employee learns a craft by working in an organisation and being taught by a master. Later, that employee may leave the firm and set up his own shop, employ new people, and teach them the craft. Jacobs argued that especially breakaways of able workers would be a source of regional development, since the entrepreneur would rarely leave the area, but stay in the place where he has the social connections. Jacobs's ideas are consistent with recent findings (some of which are reviewed below) that spinoff firms are an important engine of growth in the early phases of clusters. Experienced workers that leave a firm to start up their own firm in the same industry are a well-known mechanism for diffusion of knowledge. Historically, there have been many attempts to limit spinoffs and geographical mobility of employees in, for example, craft-based industries. Veblen (1912) colourfully described how entrepreneurs faced the death penalty if they left the area to start up somewhere else than in the cluster of glassmakers in Venice.

The transfer of routines and experience between a new firm and the previous employer of its founder has been analysed by organisational sociologists. Blueprints of the incumbent are passed on through the new firm's founders (Brittain and Freeman, 1986; Carroll, 1984; Hannan and Freeman, 1986). The relation between performance of the incumbent and the spinoff has also received attention in economics and management (e.g. Agarwal et al., 2004; Klepper, 2001, 2002a). Klepper (2001) proposes a model that combines the ideas of reproduction and inheritance with the notion of organisational routines. This notion was originally developed by Nelson and Winter (1982), assuming that firms to a large extent are governed by routines. A firm has separate routines for each of the different functions (R&D, marketing, management, etc.) and products. When a new firm is born, organisations will reproduce, because founders rely on routines they are already familiar with from their previous employment experience. The quality of these routines will determine the future success and performance of the new firm. Spinoffs may inherit more suitable routines than any other kind of start-up. This may on average enable spinoffs to outperform other start-ups.

Market growth as well as the industry and technological life cycles influence the emergence and evolution of clusters in a particular industry (Dalum et al., 2005; Klepper, 2002a). Market growth creates opportunities for additional firms to enter, and evolution of the industry opens new opportunities for market and application diversification for existing firms and spinoffs in a cluster.

Entrepreneurs may be motivated to found competing firms when the growth of their current employer indicates a substantial demand for its products and services. High growth can indicate the presence of unmet demand and great market opportunities (Romanelli and Schoonhoven, 2001). In these cases, employee learning becomes important since spinoffs can exploit the opportunities of their parents by entering with a similar product. Consequently, they either compete directly or more narrowly with one of the parent's activities (Franco and Filson, 2000).

Application diversification and innovations relate closely to the activities of parents and grow out of research undertaken there. Spinoffs will occur when employees pursue new technological discoveries made in their own firm. Wiggins (1995) and Anton and Yao (1995) add that the innovations are likely to be path-breaking and opening new sub-markets within the industry. The work of new firms and organisations is related to activities of parents and is in some cases innovative by introducing new products or services. Entrepreneurs bring knowledge and skills from their past working and educational activities that may be valuable in searching for new business areas and opportunities as well as in the daily life of running a firm (Shane, 2000). Founders are likely to bring specific knowledge about a wide range of issues to their new firm, such as customer demand, products, technologies, suppliers and competitors (Helfat and Lieberman, 2002). This may also include knowledge about how to exploit new knowledge and technological developments based on unmet supplier or customer demands or prior scientific and technical training (Roberts, 1991).

Growth in the number of firms within a particular industry in a region is by definition the most important process in the emergence of clusters. However, the emergence may stop at various stages, even when the necessary conditions seem to be present. Three key mechanisms are central determinants in the early growth phase of clusters: (1) the geographical dimension of entrepreneurial activity, (2) spinoffs from successful firms, and (3) new market opportunities. It is in the interaction between these three mechanisms that a particular industry may grow from a single firm to become a cluster. Geography obviously plays a role in the interaction between the first two determinants, but the new market opportunities relate more to the industry life cycle and the growth of markets outside the region. Location-specific externalities might emerge as the cluster evolves from a single firm to many. However, in the early phase it seems unlikely that these are attracted to a non-externality region. Internal circumstances may cause a spinoff process to stop. Jacobs (1969) describes how Rochester, New York, used to 'produce' a high number of breakaways within precision manufacturing and optical products, but turned into a company town, since the successful Eastman Kodak company started fighting breakaways and acquiring competitors. Sometimes a successful company does not 'produce' spinoffs but grows large on its own. Then a regional agglomeration within a particular industry consists of one or a few large firms.

A successful incumbent can spawn new firms that are more successful on average because their founders bring knowledge from the incumbent. Success will breed success in this respect. This spinoff mechanism is not necessarily geographic in its nature, but since most entrepreneurs found firms near their past employers, local success is more likely to generate success in the same region rather than elsewhere. The opportunities for application diversification and market diversification increase when the market grows. These three mechanisms put together can be the main driving forces of geographic clustering of economic activities, because a single successful incumbent can lead to well-performing spinoffs, which later on can spawn other new firms. To a large extent this is because entrepreneurs are likely to stay where they are, and because they are able to rely on the success they experienced while being employed at the incumbent. Furthermore, the firms in the cluster develop new applications of the initial technologies and seek out new customers. Location-specific externalities might become important as the number of firms in the cluster increases, but not in the early growth phase.

3. Methodology

The case study presented below has been developed using detailed information about the founding events and organisational backgrounds of each individual entrant in the cluster until 2003. The collection of this information has involved extensive work to trace the founders of every firm. First, we need to identify all firms that at any point in time have existed in the region. We have identified the name of the firm, the year it was founded, and the year that it closed, if this has been the case. The second step is to identify the founders and where they worked previously. The data are presented in a table in Dahl et al. (2003) with the names of all companies, year of entry, year of exit (if any), name of founders, their organisational background, and main events in the history of the firms, such as acquisition and bankruptcies. Dahl et al. (2003) also presents a more detailed description of the history of the cluster.

The information was collected from existing historical studies (Dalum, 1995; Gelsing and Brændgaard, 1988), the cluster's own organisation, NorCOM, founded in 1997, newspaper articles, information from the Internet, and a number of interviews and casual conversations with key actors.[1] An earlier version of this chapter was presented at a NorCOM Association meeting, where many key actors were present. Many of these have been employed at the early firms and have experienced the growth of the cluster. The comments from key industry and university participants at this meeting have been a valuable source. All information has been double-checked to make sure that the information is accurate using multiple different sources.

4. The emergence of a wireless communications cluster

The case to be studied in detail is the emergence of a cluster of high-technology-based firms in the field of wireless communications (NorCOM) in the region of North Jutland. Defined narrowly, it consists of around 50 firms.

Simon Petersen founded S.P. Radio in 1948 as a radio and later TV manufacturer in Aalborg. This industry faced increasing competitive pressure in the early 1960s, which fuelled a shakeout. Meanwhile, Simon Petersen had noticed opportunities in the market for maritime radio communications equipment for small vessels. Consequently, S.P. Radio managed to diversify into maritime communications from the mid-1960s and became one of three world leaders within half a decade up to the early 1970s. In 1966, the company had reached 150 employees.

In 1973, the first spinoff firm, Dancom, was founded by three engineers from S.P. Radio. One of them had been the head of R&D at S.P. Radio for four years. Dancom was active in the same markets as S.P. Radio, producing maritime communications equipment. Two former Dancom engineers founded the second spinoff, Shipmate, in 1977. Their first product was a radio phone for maritime use, which competed directly with both Dancom and S.P. Radio. Their breakthrough came in 1980, after they developed a satellite navigation system (application diversification). At that time, the firm only had one employee besides the two founders. Shipmate successfully developed and produced radiophones and navigation equipment for maritime use. Five years later, the firm had reached 200 employees. Dancom went into financial difficulties in 1980. The firm started to explore the possibilities of using the firm's capabilities and advantages from maritime communications in an emerging market of onshore personal communications (mobile phones). Shortly after, the firm started activities in producing personal mobile phones and changed its name from Dancom to Dancall Radio.

In the early 1980s, some of the cluster firms and spinoffs diversified into an emerging technologically related area of personal mobile communications equipment. This offered new opportunities as the mobile communications market opened with the introduction of the common Nordic standard for mobile telephony (NMT). When the market boomed during the 1980s, these firms were among the world-leading producers of phones for this network. They were able to use the inherited and developed strong capabilities from the maritime radio communications. In 1985, Shipmate expanded into mobile phones with the new activities placed in a separate division, Cetelco. After two years, Cetelco had 25 engineers working with R&D. They developed and produced mobile phones for several European and East Asian markets.

The cluster firms experienced high growth because of the success of local firms that diversified into mobile telephony, and the North Jutland region became visible as a strong region in this industry. However, in the last half of the 1980s the market changed because of rapid technological development. The small firms were facing an international market with high development costs and production capacity demands, and intense price competition. The industry life cycle and market growth influenced the cluster significantly, but they also opened up new opportunities for diversification as the market and technology expanded further.

In 1987, seven experienced engineers from Dancall founded T-Com. The engineers disagreed with Dancall's overall market strategy and decided that they could do it better themselves. T-Com's strategy was to develop mobile phones just like its parent company, but differed by only focusing on R&D as a subcontractor. Other companies would then produce and market the phones under their own brands. In the same year, they developed their first mobile phone. In the last half of the 1980s, there were a total of 15 firms in the industry in Northern Denmark. One of the entrant firms had been closed, but new firms had continued to enter. The majority of these were spinoffs.

The common European standard for mobile telephony (GSM) was implemented as a new standard in 1992. The success of the NMT standard inspired the European telecommunications operators to create a common European system based on digital technology. A race began between the leading producers in the world to be the first to be able to produce a complete terminal for this network. The challenge of building a GSM mobile phone was seen to be a major economical and technical challenge for the mobile communications firms, since it was based on new digital technology. To cope with this, the two competitors Dancall and Cetelco formed a pre-competitive joint venture company, DC Development, with the purpose of building the basic modules of a mobile phone in close cooperation with Aalborg University.

DC Development was founded in 1988 and located at a new science park, NOVI, close to Aalborg University. DC Development succeeded in development of basic modules, and the parent companies were among the first in the world to produce a GSM mobile phone in 1992. This technological achievement increased the international visibility of the cluster and strengthened the region's reputation in wireless communications technologies. However, the new standard changed the market considerably as large multinationals continuously entered the promising GSM market. The consequence was increased competition, falling prices, rapid development and increased demand for volume production. The high development costs of GSM phones put Dancall and Cetelco into severe financial problems in the early 1990s, because they did not have enough financial backup to harvest their discoveries in this competitive market.

As a consequence, Cetelco was gradually taken over by Hagenuk (Germany) in 1988–90. Dancall had more than 600 employees in early 1993. But they also had a troubled period. The newly produced GSM phone was too expensive compared to competitors and at the same time the export of NMT phones suffered from declining markets as the new GSM market grew. Consequently, Dancall had severe financial problems and was reconstructed again and sold to Amstrad (United Kingdom). During the next couple of years, the firm gained momentum and grew to become larger than before the reconstruction.

During the history of the cluster, Dancall and Cetelco have been the parents of numerous spinoffs. Many of these are clear examples of the market and application diversification that has been dominating the mobile phones industry. An example is RTX Telecom founded in 1993 by three experienced engineers from Maxon and four from Cortech. They founded the firm with the strategy of doing R&D for other firms. RTX based their designs on chipsets from National Semiconductor who had previously worked with Dancall and needed the knowledge on the future development of wireless devices to design their chipsets, and RTX needed the chipsets. RTX grew from seven employees in 1994 to more than 200 in 2003. Another example is ATL Research founded by engineers from Cetelco in 1996. While working at Cetelco, they often got enquiries from other firms in the industry who wanted to buy development aid for mobile phones. But it was not possible to follow this potential market within Cetelco. ATL cooperated with several chipset manufactures until Texas Instruments acquired the firm in 1999.

In the late 1990s, the mobile communications industry had high growth rates and the large multinational players in the industry were increasingly looking for new regions for their activities in order to access local pools of development engineers. As a result, many multinationals were attracted to the region.

The success of the first company in the NorCOM cluster, S.P. Radio, in the 1960s and 1970s as one of the world's leading producers of maritime communications equipment gave the employees who were the founders of the first spinoffs in the 1970s the relevant capabilities and routines to become successful themselves. Market opportunities, and market and application diversification were important in the evolution of the cluster. Some of the first spinoffs, Dancall and Cetelco, successfully diversified into mobile phones by using their experiences gained by the founders, while they worked at the parent, S.P. Radio. The growth of the mobile communications industry opened new opportunities that, coupled with the initial success of these firms, fuelled an intense formation of new spinoffs as the market grew, which developed the cluster as it grew from one to 50 firms during three decades. Application diversification and evolution of the industry life cycle have widened the knowledge base. The core competence is still wireless communications, but it has expanded from maritime communications to land-based radio, mobile phones, cordless phones, and many other wireless technologies. The maritime communications also led to radio-based navigation, satellite navigation, and fleet management systems.

5. The role of spinoffs in the early growth phase

We have collected detailed information on every firm active in the cluster since the 1960s. They have been divided into different groups based on their prior experience: diversifying entrants (firms entering new or established markets unknown to them), parent spinoffs (new entities founded by established companies), joint ventures, entrepreneurial spinoffs (firms founded by experienced persons with previous employment in incumbent firms in the industry) and inexperienced entrants (founded by persons with no previous employment in the industry). The evolution of the population of firms, based on types of entry is shown in Figure 9.1. We have intentionally left out the two joint venture firms to keep the graph visible.

Figure 9.1 shows that the entrepreneurial spinoff firms represented a large share of the increase in the population. They were a driver of cluster growth during the entire period. They represent the largest group of firms. The total population has grown to more than 50 firms and the number of exits is low. Several firms have, however, been rescued from exit on the verge of bankruptcy after being taken over, typically by multinational companies.

Figure 9.1 shows that the second firm entered in 1973 as a spinoff from S.P. Radio. A decade later the number of firms had increased to six of which three were spinoffs, two were diversifiers (including S.P Radio) and a single inexperienced entrant. The introduction of mobile communications in Denmark in 1981 and the subsequent market growth influenced the emergence of the cluster. The success of Dancall and Cetelco and their growth in employment resulted in several spinoffs. From 1984 to 1990 the population of firms increased from 6 to 16. Spinoffs accounted for half of these. From 1990 to 2003, the population increased from 16 to 51. In the early 1990s, several inexperienced start-ups entered the cluster and parent spinoffs entered in the late 1990s. However, the entrepreneurial spinoffs hold the largest share of the population of firms.

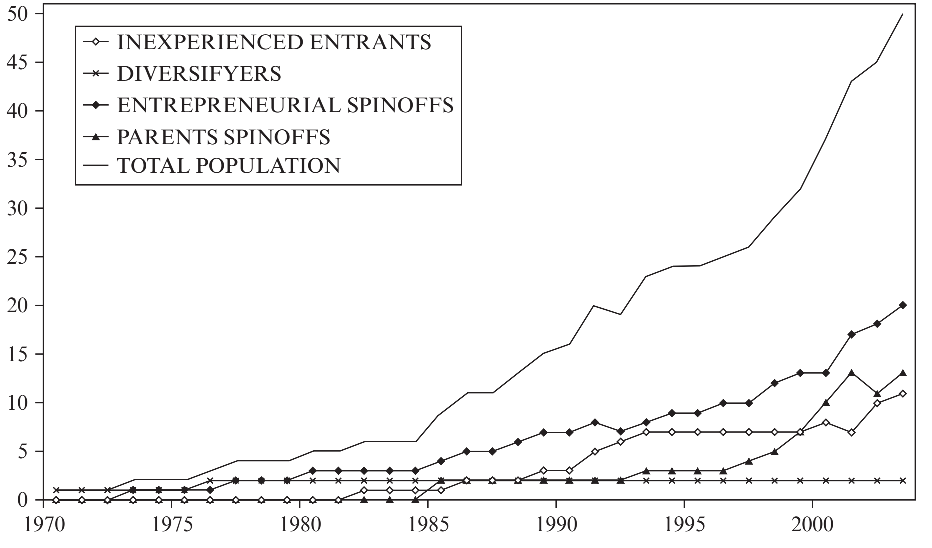


Figure 9.1 Evolution in the population of firms in NorCOM by main category (1970–2003)

The growth in the number of parent spinoffs since 1997 represents the increasing international visibility of the cluster. At the introduction of the GSM system, where local firms were at the frontier of the technology from the beginning, the international focus on North Jutland's wireless industry has increased the number of parent spinoffs from three in 1996 to 15 in 2001. Before 1997, there were mainly local or national based players behind the parent spinoffs. Later, it was foreign companies, such as L.M. Ericsson (Sweden), Texas Instruments (United States), Motorola (United States), Siemens (Germany), and National Semiconductors (United States), which entered through acquisition of or investments in already established local firms. Multinationals such as Nokia (Finland), Lucent (United States), Analog Devices (United States), and Infineon (Germany) located activities in the region through green field investments.

The three large early firms, S.P. Radio, Dancom/Dancall and Shipmate/Cetelco have been the source of many spinoffs. In total, these three have been the source of 12 new spinoffs. Employees coming from Dancom/Dancall founded eight of these firms. Many of the founders had also been employed at S.P. Radio before going to Dancom/Dancall. This illustrates how important the three firms have been as a training ground for founders of new firms. The spinoffs gained the relevant capabilities, routines, and industry specific knowledge that helped them to become successful themselves.

6. Discussion

A successful company is not a sufficient condition for the emergence of a cluster. Two regions with similar initial conditions may develop quite differently. The Struer region in Western Jutland is an example of a region with a very successful company that never 'grew' into a cluster. It is a historical parallel to the NorCOM cluster in North Jutland. Initially they had some common features, but their development trajectories ended up being significantly different. The first firm, Bang & Olufsen (B&O), grew large and became very successful in the manufacturing of high-end radios and television sets, but it only produced a few spinoffs, and Struer remained a 'one-company' town.

Peter Bang and Svend Olufsen founded B&O in 1925 in the small rural town of Struer. Olufsen was born there, while Bang was born in Copenhagen. Olufsen's family hosted B&O in the first years at the family farm. From the late 1920s, B&O grew considerably and was already in the 1930s an established radio manufacturer and experienced strong employment growth. The manufacturing activity was contained within one firm for many decades and a cluster never emerged in Struer. B&O experienced several spinoffs, but only two of these stayed in the region. The Struer development, which had gained significant international reputation already in the 1950s, never resulted in the creation of a cluster and had the character of single company concentration.

The emergence of the wireless communications cluster in North Jutland can hardly be explained by various location-specific externalities. S.P. Radio was the only company in the industry in the region for many years. It also seems unlikely that there was a unique local culture, climate or natural resources that, somehow, were ideal and designed for wireless communications technologies. Simon Petersen did not decide to found S.P. Radio in North Jutland to tap into tacit knowledge or to benefit from labour market pooling, knowledge externalities or specialised suppliers. Likewise, B&O located in Olufsen's hometown at his parents' farm and did not benefit from concentration of resources. Accordingly, the founding of both firms can be characterised as chance events driven by sociological factors rather than economic choices of optimal locations.

These new firms had an effect on regional industry structure, but initially it was a small effect. It did not start the path-dependent agglomerative processes that deterministically lead to a cluster. B&O was successful and experienced a large growth in employment. However, this success did not create a cluster. B&O managed to contain its radio and TV activity within one firm. During the history of the firm, it produced a few spinoffs, but these decided to locate outside Western Jutland, except for two in the 1980s. S.P. Radio existed for many years before it diversified into maritime radio communications. However, a growing number of firms in a cluster are by definition the most important process in its emergence. The second firm entered as a spinoff from S.P. Radio and in the early 1980s the 'cluster' consisted of six firms of which three were spinoffs. In the early 1990s, there were 16 firms working with wireless communications or related technologies, and spinoffs still accounted for half of the population. The strong employment growth in North Jutland occurred in the 1980s and 1990s, while the strong employment growth of B&O was already envisaged in the interwar period. However, the spinoff process pushed the early growth phase of the cluster in North Jutland, while Struer became a one-company town. Thus spinoffs are a key process in the emergence of clusters and it cannot be explained by interaction between parts of Porter's diamond that are present in a location because of historical circumstances. However, emergence of a cluster might suddenly stop at various stages, even though the necessary conditions seem to be present and spinoffs emerge. It is not a deterministic process.

In the early growth phase of the wireless communications cluster the role of geography was mainly confined to the limited mobility of start-ups. It was very hard to attract labour to the region in the early phase. An engineer who actually moved from Storno, a wireless communications firm in Copenhagen, and later founded a spinoff, confirms this. He claims that it was not attractive to move to North Jutland to get a job at Dancall in the early 1980s, because there were only a few firms to work for. However, as the market grew and additional firms entered the industry in the region and the pool of skilled labour increased, the role of geography changed. The increased market opportunities especially in mobile communications and the growing number of firms and labour also increased the breeding ground for spinoffs. It also opened the possibility for labour mobility between firms and subsequently the creation of social networks between employees working in various firms. In the 1980s and 1990s, the university had an effect through its growing supply of qualified labour and basic research in the wireless technologies. Later, firms were attracted to the region and entry by parent spinoffs. Thus, traditional cluster effects, such as knowledge spillovers, labour market pooling, and attraction of firms, can be found in the cluster at a later stage.

The importance of entrepreneurial spinoffs is known from other studies of industry evolution. These studies focus on the role of entrepreneurial spinoffs in the evolution of particular industries that also created well-known clusters, such as semiconductors in Silicon Valley (Moore and Davis, 2001), wireless communications in San Diego (Simard and West, 2003), automobiles in Detroit (Klepper, 2002a), and tyres in Akron (Buenstorf and Klepper, 2005).

Many studies of evolution of the semiconductor industry in Silicon Valley have shown that entrepreneurial spinoffs and employee mobility have been important factors.

Brittain and Freeman (1986) and Moore and Davis (2001) argue that entrepreneurial spinoffs were the main engine behind the rapid growth in Silicon Valley. In the early years, the semiconductor firms faced novel organisational challenges in aligning goals and designing the organisational structure that could establish and reach the technological demands of the industry. Working for incumbent firms was the best way to learn how to tackle those challenges. The most well known case is the many firms that spun off from Fairchild Semiconductor, often called the 'Fairchildren', such as Intel and AMD.[2] In general, engineers left established incumbents and started new firms that produced the capital goods and materials needed for semiconductor design and manufacturing.

Simard and West (2003) analyse the emergence of the wireless communications cluster in San Diego. Linkabit was founded in 1968 initially doing general military consulting, but later it diversified into secure military communications. In the 1980s, Linkabit spawned four spinoffs, Radyne, Tiernan, PCSI and Qualcomm. The latter was founded in 1985 and the founders saw an opportunity to adapt military communications technologies for civilian use. The timing of the market diversification coincided with the take-off in the mobile communications market. Qualcomm successfully developed CDMA technology that was approved as one of three standards for digital mobile communications in US in 1990. Qualcomm experienced high growth during the 1990s. However, the increased market opportunities were coupled with a spinoff process from the first few firms in the emerging cluster. During the 1990s, the cluster became visible and in the late 1990s the major players in the industry was attracted to the region, such as Nokia, Ericsson, Siemens, Samsung, and Sony.

Studies of the evolution of the US automobile industry and the emergence of the cluster in Detroit also emphasise the importance of spinoffs (Carroll et al., 1996; Klepper, 2002b). The industry was characterised by a high rate of entry with more than 500 firms entering in its first 20 years. Eventually, the industry evolved to be an oligopoly dominated by three relatively late entering Detroit firms, Ford, General Motors and Chrysler. These three were all related to the first successful firm in the industry, Olds Motor Works, which was located in Detroit (Klepper, 2002b). These firms spawned 22 spinoffs. This growth in the number of firms and the emergence of a cluster were supported by the high growth in the market for automobiles. By the 1930s, Detroit firms dominated the industry.

Buenstorf and Klepper (2005) analysed the evolution of the US tyre industry. They found that the location and background of the entrants caused the industry to become regionally concentrated around Akron, Ohio. The tyre industry evolved to be an oligopoly dominated by Goodyear, Goodrich, Firestone, and U.S. Rubber. These four firms accounted for more than 70 per cent of the market in the 1930s. Goodrich, Goodyear and Firestone were all located in Akron. The Ohio firms were distinctly successful and the production of tyres and inner tubes in Ohio accounted for two-thirds of total US production in 1935. The Akron tyre cluster emerged through a spinoff process as the market grew. The leading Akron firms spawned the most spinoffs and the performance of these was very much related to the performance of their parent. From a few early Akron firms related to the successful Goodrich, the Akron tyre cluster grew primarily through a spinoff process rather than through agglomeration economies (Buenstorf and Klepper, 2005).

The evolution of the semiconductor industry in Silicon Valley, the wireless communications industry in San Diego, the automobile industry in Detroit and the tyre industry in Akron illustrates the importance of the spinoff process in the early growth phase of clusters. The very detailed studies of the organisational background of entrants highlight the mix of a successful firm spawning successful spinoffs that stay in the region and are able to exploit new market opportunities. The role of geography is limited in the early phase and several regions might have had similar initial conditions. Likewise there do not seem to have been any given advantages in the regions; for example Detroit had historically been known for its production of flour and copper mines (Jacobs, 1969), North Jutland was dominated by agriculture, fishing, ship building, tobacco, textiles and construction materials, and military aerospace was San Diego's major manufacturing industry from the 1930s through the 1980s (Simard and West, 2003). New market opportunities seem to be important in relation to the growth of clusters. The opportunities and market growth of mobile communications supported the emergence of the wireless communications clusters in North Jutland and San Diego.

7. Conclusion

The dominant theories on industrial clusters focus more on explaining ex post dynamics of clusters than the emergence of clusters. As a result many case studies often contain unique and colourful explanations as to why a particular industry became clustered in a particular location. One of the risks in these studies is that the intriguing chain of events that eventually lead to clusters clouds more general processes. Consequently, geography appears to be a very important factor in the emergence of clusters, because of the founder of the first firm's decision to locate in a particular region. In addition firms are shaping the regional environment through their demand for labour and knowledge and they are also affected by the existing structure of the region. However, we argue that the role of geography in the emergence of clusters is mainly limited to the geographical dimension of entrepreneurial activity, that is, entrepreneurs tend to found their firm where they live. The founding of the first firm can be characterised as a chance event, but a successful firm becomes a training ground for spinoffs that are able exploit new market opportunities.

In the case of wireless communications in North Jutland, the success of S.P. Radio helped the spinoffs to gain routines to become successful themselves. The entry of these firms was a key process when the cluster grew from one firm to many. Spinoffs remained the dominating share of firms from the mid-1970s and onwards, but were particularly important in the early growth phase when there were no apparent location-specific advantages. The move into the emerging market for mobile phones in the early 1980s was important for further growth. The new market opportunities in the first and second generation of mobile phones and growth in demand sustained the employment growth in Dancall and Cetelco, and paved the way for additional spinoffs. This demand was not local, but initially Danish. During the 1980s the NMT standard was introduced in many European and Middle East countries in various versions and the cluster firms exported to these countries. The worldwide increase in demand for mobile phones also helped the emergence of the wireless communications cluster in San Diego. Spinoffs from the early firms, Linkabit, Qualcomm, PCSI, benefited from the market conditions and their inherited competences. Likewise, the clusters in Silicon Valley, Detroit and Akron emerged over the industry life cycle. Thus, the new market opportunities and increased demand facilitated their emergence. However, the process of clustering was not deterministic and could have been stopped at various stages, when technologies changed, new products failed, competition increased or the market changed. In the case of North Jutland, two of the seedbeds for many spinoffs, Dancall and Cetelco, experienced severe financial problems on several occasions and were taken over by foreign firms. If these had been closed down at an early stage it might had hampered the emerging cluster.

Local conditions do not appear to have been important in the emerging phase of the cluster in North Jutland. Later on, in the 1980s and 1990s, other factors supported the growth of the cluster, such as the university through its growing supply of qualified labour and basic research in wireless technologies. Thus, the role of geography changed after the clusters emerged. Studies of knowledge diffusion between engineers across firm boundaries in the cluster in 2001 reveal that they shared knowledge (Dahl and Pedersen, 2004), but these spillovers could not take place without several firms that emerged via the spinoff process.

The main finding in this chapter is that spinoffs have been a central mechanism in the early growth phase of the wireless communications cluster in North Jutland. This result is supplemented by the findings in semiconductors in Silicon Valley, wireless communications in San Diego, automobiles in Detroit, and tyres in Akron. Without the spinoff process, the NorCOM cluster would never have become a cluster, but only a single firm like B&O in Struer. This explanation for the emergence of clusters removes some of the uniqueness and historical and cultural specificity of clusters. Spinoffs are a key determinant when a cluster grows from a single firm to many firms. However, there are several unanswered questions concerning the spinoff process and the process of founding a firm. Why do spinoffs occur in some places and not in others? What triggers the separation process? What type of positions and functions do future founders have in the parent firm? Which types of routine are inherited? What is the link between the industry life cycle and spinoff process? Answers to these questions can give us much better knowledge about the process of spinning off and the emergence of clusters. Future studies of the emergence of clusters need to dig deep and study the organisational background of founders, the evolution of the industry and market opportunities.

Acknowledgements

The authors are grateful for comments received from Ron Boschma, Magnus Holmen, Steven Klepper, Olav Sorenson, Niels-Christian Gjerrild, Jens Hansen, Henrik Langkilde and Erik Rauff.

Notes

1. Such as Erik Rauff (co-founder of Shipmate, former employee of S.P. Radio and Dancom/Dancall), Henrik Langkilde (CEO of Dancall in the 1980s), Niels-Christian Gjerrild (co-founder of T-Com, CEO of Dansk DECT/Ericsson, former employee of Dancom/Dancall), Jens Hansen (co-founder of RTX, former employee of Dancom/Dancall and Cortech) and many others, who have been employed at the early firms and/or been part of founding teams.
2. Fairchild Semiconductor was a spinoff from Shockley Transistor. The latter was founded in 1955 by the co-inventor of the transistor, William Shockley, who previously worked for Bell Telephone Laboratories.

References

Agarwal, R., R. Echambadi, A.M. Franco and M.B. Sarkar (2004), ‘Knowledge transfer through inheritance: spin-out generation, development and survival’, *Academy of Management Journal*, **47** (4), 501–22.

Anton, J. J. and D.A. Yao (1995), 'Start-ups, spin-offs, and internal projects', *Journal of Law, Economics, and Organizations*, **11**, 362–78.
Arthur, W.B. (1990), 'Silicon valley locational clusters: when do increasing returns imply monopoly?', *Mathematical Social Sciences*, **19**, 235–51.
Arthur, W.B. (1994), *Increasing Returns and Path Dependence in the Economy*, Stanford: University of Michigan Press.
Boschma, R. and G.A. v. d. Knaap (1999), 'New high-tech industries and windows of locational opportunity: the role of labour markets and knowledge institutions during the industrial era', *Geografiska Annaler*, **81B** (2), 73–89.
Brittain, J.W. and J. Freeman (1986), 'Entrepreneurship in the semiconductor industry', Mimeo.
Buenstorf, G. and S. Klepper (2005), 'Heritage and agglomeration: the Akron tire cluster revisited', Mimeo.
Carroll, G.R. (1984), 'Organizational ecology', *Annual Review of Sociology*, **10**, 71–93.
Carroll, G.R., L.S. Bigelow, M.-D.L. Seidel and L.B. Tsai (1996), 'The fates of *de novo* and *de alio* producers in the American automobile industry 1885–1981', *Strategic Management Journal*, **17** (Summer Special Issue), 117–37.
Dahl, M.S. and C.Ø.R. Pedersen (2004), 'Knowledge flows through informal contacts in industrial clusters: myth or reality?', *Research Policy*, **33**, 1673–86.
Dahl, M.S., C.Ø.R. Pedersen and B. Dalum (2003), 'Entry by spinoff in a high-tech cluster', *DRUID Working Paper Series*, 2003–11.
Dalum, B. (1995), 'Local and global linkages: the radiocommunications cluster in Northern Denmark', *Journal of Industry Studies*, **2** (2), 89–109.
Dalum, B., C.Ø.R. Pedersen and G. Villumsen (2005), 'Technological life cycles: lessons from a cluster facing disruption', *European Urban and Regional Studies*, **12** (3), 229–46.
Franco, A.M. and D. Filson (2000), 'Knowledge diffusion through employee mobility', *Research Department Staff Report*, Federal Reserve Bank of Minneapolis.
Gelsing, L. and A. Brændgaard (1988), *Elektronikindustrielle Miljøer i Nordjylland* [Electronics Industry Environments in North Jutland], Aalborg: Institut for Produktion, Aalborg Universitets Center.
Hannan, M.T. and J. Freeman (1986), 'Where do organizational forms come from?', *Sociological Forum*, **1**, 50–57.
Helfat, C.E. and M.B. Lieberman (2002), 'The birth of capabilities: market entry and the importance of pre-history', *Industrial and Corporate Change*, **11** (4), 725–60.
Jacobs, J. (1969), *The Economy of Cities*, New York: Vintage Books.
Klepper, S. (2001), 'Employee startups in high tech industries', *Industrial and Corporate Change*, **10**, 639–74.
Klepper, S. (2002a), 'The capabilities of new firms and the evolution of the US automobile industry', *Industrial and Corporate Change*, **11** (4), 645–66.
Klepper, S. (2002b), 'The evolution of the U.S. automobile industry and Detroit as its capital', Mimeo, Carnegie Mellon University.
Marshall, A. (1920), *Principles of Economics*, London: Macmillan.
Maskell, P., H. Eskelinen, I. Hannibalsson, A. Malmberg and E. Vatne (1998), *Competitiveness, Localised Learning and Regional Development – Specialisation and Prosperity in Small Open Economies*, London: Routledge.
Moore, G. and K. Davis (2001), *Learning the Silicon Valley Way*, Stanford: Stanford Institute for Economic Policy Research.
Nelson, R.R. and S. Winter (1982), *An Evolutionary Theory of Economic Change*, Cambridge, MA: Harvard University Press.
Porter, M.E. (1998), *On Competition*, Boston, MA: Harvard Business School Press.
Roberts, E.B. (1991), *Entrepreneurs in High Technology*, New York: Oxford University Press.
Romanelli, E. and C.B. Schoonhoven (2001), 'The local origins of new firms', in C.B. Schoonhoven and E. Romanelli (eds), *The Entrepreneurship Dynamic: Origins of Entrepreneurship and the Evolution of Industries*, Stanford, CA: Stanford University Press.
Shane, S. (2000), 'Prior knowledge and the discovery of entrepreneurial activities', *Organization Science*, **11**, 448–69.
Simard, C. and J. West (2003), 'The role of founder ties in the formation of San Diego's "wireless valley"', DRUID Summer Conference 2003, Copenhagen.
Sorenson, O. (2003), 'Social networks and industrial geography', *Journal of Evolutionary Economics*, **13** (5), 513–27.
Sorenson, O. and P.G. Audia (2000), 'The social structure of entrepreneurial activity: geographical concentration of footwear production in the U.S., 1940–1989', *American Journal of Sociology*, **106**, 424–62.
Sorenson, O. and T.E. Stuart (2001), 'Syndication networks and the spatial distribution of venture capital investments', *American Journal of Sociology*, **106**, 1546–88.
Stinchcombe, A. (1965), 'Social structure and organizations', in J.G. March (ed.), *Handbook of Organizations*, Chicago: Rand McNally, pp. 153–93.

Storper, M. (1997), *The Regional World: Territorial Development in a Global Economy*, New York: The Guilford Press.
Storper, M. and R. Walker (1989), *The Capitalist Imperative: Territory, Technology, and Industrial Growth*, Oxford: Basil Blackwell.
van der Linde, C. (2003), 'The demography of clusters – findings from the cluster meta-study', in J. Bröcker, D. Dohse and R. Soltwedel (eds), *Innovation Clusters and Interregional Competition*, Berlin, Heidelberg, New York: Springer Verlag, pp. 130–49.
Veblen, T. (1912), *The Theory of the Leisure Class: An Economic Study of Institutions*, New York: Macmillan.
Wiggins, S.N. (1995), 'Entrepreneurial enterprises, endogenous ownership, and the limits to firm size', *Economic Inquiry*, **33**, 54–69.

10 A social-evolutionary perspective on regional clusters*

Udo Staber

1. Introduction

The observation that something like an 'evolutionary turn' may be taking place in economic geography (e.g. Boschma and Frenken, 2006; Boschma and Martin, 2007; Grabher, 2009) does not in itself suggest that other theoretical approaches to questions of economic geography have lost momentum. This is most apparent in research on regional clusters.[1] By theorizing from a range of perspectives to highlight issues related to institutions, technology, power, knowledge, culture, and so forth, researchers have made significant contributions to our understanding of clusters and the processes that drive them. However, the multitude of theoretical approaches in use has also fragmented the field, producing a plethora of models and many difficult-to-measure concepts that, when seen in isolation, limit the generalizability and applicability of research findings. Theoretical fragmentation is probably unavoidable in research on constructs as fuzzy as clusters and related concepts such as flexible specialization, collective learning, and identity. But there is also the need for explanations that combine generality and parsimony with rich descriptions of local phenomena. The social-evolutionary approach, intended as a general framework for linking disparate parts and perspectives on developmental aspects of clusters, may meet that need in several respects.

Many evolutionary accounts of clusters lack precision. They don't provide a clear answer to some of the fundamental questions raised in evolutionary economic geography. For example, what does it mean to say that a cluster has evolved? Has the cluster merely changed, or has it moved toward greater adaptive complexity? How did a particular cluster arise where and when it did? At which point does a cluster enter a new phase in its life-cycle? What exactly is involved when a cluster reaches a threshold (e.g. towards chaos) or faces critical values (e.g. initiating system shifts)? What are the mechanisms involved in local knowledge flows, and why might their impact differ across local settings? Under what conditions and how do institutional and organizational rules evolve over time? Routines and competencies in knowledge spillover or collective learning do not reproduce themselves automatically or spontaneously. They are acted on by individuals who replicate, reproduce, or transform them, for reasons that must be explained rather than assumed. Not all agentic interventions are successful; many are harmful, yet they may persist. In short, we need a better understanding of why the clusters we observe in different places are the way they are, rather than should be.

To this end, I propose an evolutionary idea-based view of how human agency is implicated in the ongoing social construction of a cluster. I do not develop here a full-blown evolutionary theory of agency, nor am I presenting a set of testable hypotheses. Instead, I present the fundamentals of a conceptual scheme that is intended as a 'sensitizing' framework, to highlight the role of ideas (cognitive and normative beliefs) and their relationship to human agency in the evolution of clusters. Drawing on those parts of evolutionary theory that consider knowledge a process and outcome of social construction is in keeping with the emphasis in the cluster literature on knowledge as a core asset (Tallman et al., 2004) and on learning as a key process of clusters (Malmberg and Maskell, 2002). Although it would seem that the ideas underlying knowledge are central to any knowledge-based and learning-based approach to clusters, ideas and the interactions between them have been largely absent from most accounts of clusters, including those emphasizing interpersonal relations at the micro-level (Sunley, 2008, p. 10). The evolutionary framework suggested here focuses on ideas as the unit of selection, the dynamic variation of populations of ideas, and the mechanisms by which ideas travel across time and space.

The interest in ideas and agency complements the insights from other evolutionary perspectives, such as complexity theory and population ecology. Scholars working from these perspectives have tended to relegate the human actor to a residual category and to describe the diffusion of ideas as a simple contagion process. This is problematic if it condemns us to sterility of theorizing. We need to know, for example, what kind of information flows through networks before we can say how much learning has taken place or what the purposes of learning are for the actors involved. Knowledge and ideas are central to understanding human agency – intentional and nonintentional – in the social sciences. At its best, the social-evolutionary approach proposed here provides a framework for understanding how ideas are implicated in the general Darwinian processes of variation, selection, and retention that underlie human behavior. Although this view draws on basic principles from evolutionary biology, it decidedly does not involve biological reductionism (Campbell, 1965; Dennett, 1995; Hodgson, 2007; Richerson and Boyd, 2005). It places individuals within a field of social relationships that is continually reconstituted through actions. Agency is revealed in the *constitution* of meaning as a cognitive achievement as well as the *construction* of meaning as a social achievement.

I begin this chapter with a brief outline of how different variants of evolutionary theory have addressed some of the central concerns in cluster 'theory'. I then discuss the basic components of the social-evolutionary perspective. My main conclusion is that evolutionary research on clusters is thriving, but it would benefit from a more systematic concern for idea-based processes and social-constructionist mechanisms linking actors and actions across levels of analysis.

2. Variants of evolutionary thinking

As a central place for the creation and reproduction of organizational forms (Hannan et al., 2007), behavioral routines (Nelson and Winter, 1982), technologies (Rigby and Essletzbichler, 1997), and intangible assets (DiTommaso et al., 2004) such as trust (Lorenzen, 2002) and identity (Romanelli and Khessina, 2005), clusters are situated at the interface of a multitude of forces and levels of action. These include the accumulated interpretations of competencies carried by cluster actors, the competitive and institutional rules in the community, and the differential reproduction of strategic options. The organizational forms of clusters develop in a continuous process of adaptation to environmental conditions that co-evolve in response to adaptation behaviors, sometimes by design and sometimes by chance, and often with unpredictable outcomes. The success of cluster firms and the cluster as a whole depends, from an evolutionary perspective, on the ability of actors to create new variations and to retain adaptive variations.[2] As long as the organizational forms of clusters enable variation and retention and are subject to selection forces, evolution can take place.

Evolutionary theorists are well aware that the processes and mechanisms by which evolution unfolds in the social domain can differ significantly across and within types of social system, as opposed to the more 'single-mindedness' of genetic evolution. These differences, however, do not undermine the view that ongoing change is Darwinian in that the general evolutionary processes of variation, selection, and retention occur in all units and at all levels of social systems (Dennett, 1995; Hodgson and Knudsen, 2006). These generic processes subsume additional social processes, such as learning and institutionalization, which may play out differently in different contexts, depending on the details of the mechanisms involved (Staber, 2009). The challenge is to discover exactly how these processes operate in the socio-economic setting of a given cluster.

Evolutionary theory is not a monolithic bloc but incorporates a variety of more or less distinct approaches that cluster researchers have found useful for exploring issues related to variety, path dependence, uneven development, life-cycle, competitiveness, entrepreneurship, and so forth. Scholars have borrowed evolutionary concepts from organization science, particularly population and community ecology, and from institutional theory and complexity theory. The ecological perspective proposes that change be studied at the level of organizational populations, defined as collectivities whose members draw on the same or similar resources (Hannan et al., 2007). To this end, they have employed Darwinian evolutionary models to explain the evolution of populations in terms of vital rates, such as rates of organizational foundings and disbandings. The extension of this work to broader communities of interacting populations is useful for cluster research in that it allows the incorporation of organizations located outside the focal cluster as a potential source of new variation. The study of populations occupying distinct locations in geographical space makes it possible to examine questions related to the constitution and construction of cluster boundaries, such as the role of competition and mutualism in reproducing boundaries. It thus speaks to the central concern in cluster research about the geographic reach of resource linkages.

At the organizational level, evolutionary analysis proceeds on the premise that organizations are not unitary entities but more or less well coordinated bundles of routines and competencies. Routines may either substitute or complement each other. In cases where routines 'compete', they may drain vital resources from the organization, as when firms that invest too much in exploiting existing resources do not have sufficient capacity left to explore new opportunities (March, 1991). The concern for competition and mutualism draws attention to the ambivalent effects of heterogeneity within firms and clusters. By increasing cognitive distance between cluster firms, internal cluster heterogeneity may limit social integration and knowledge sharing; but this may also be a source of variation, stimulating innovation.

Institutional theory is often employed to explain how organizations seek strategies and structures that fit the social context in which resource selection decisions are made (Scott, 2001). Institutionalization of structures and behaviors is seen as both an outcome (drawing attention to stability and retention) and an activity (drawing attention to change and variation). The institutional approach has been found useful by economic geographers interested in questions of coordination, domination, and legitimation. A cluster is said to be 'institutionally rich' if it includes numerous institutions offering a variety of resources, high levels of interaction among these institutions, collective structures that emerge as a result of this interaction, and a mutual awareness of the necessity of developing common objectives and unified action (Sydow and Staber, 2002). The evolutionary interest in institutions is not so much directed at understanding institutions as formal organizations, but is motivated by a concern for the meanings and social disputes that create, transform, and reproduce specific institutional arrangements.

Complexity theory has also provided a number of useful insights for the evolutionary analysis of clusters (Martin and Sunley, 2007). It views clusters as open and complex adaptive systems, driven by non-linear, self-organizing, and co-evolutionary dynamics, with emergent properties based on spontaneous interaction among system components (McKelvey, 1997). One of the arguments taken from this theory is that clusters are most robust, innovative, and adaptive if they exist at the 'edge of chaos', a position that balances order and chaos, control and change, and cooperation and competition. The evolving balance between robustness and fragility, between staying the same and changing into something else, has yet to be explored more fully in research on clusters. The most successful clusters are not necessarily, as is often argued in this literature, those that are most innovative, in the sense that they generate a continuous stream of new ideas. Constant change and continuous mutations may make it impossible for the actors to apply new discoveries, to the extent that the exploitation of discoveries requires some degree of stability to retain useful knowledge, create common understanding, and establish accountability. Successful clusters may be flexible, but they also contain stabilization dynamics without which the benefits obtained from selected variations would dissipate and change would be random.

The principle of self-organization states that the cluster 'tunes itself' spontaneously towards an order, without an overall plan and central direction. However, what exactly this means in particular instances is left unclear especially in analyses that remain stuck at a normative level, with arguments about complex systems being 'effective' if they are 'somewhere in the middle' between too much and too little complexity, or if they provide an 'optimal mix' of homogeneity and heterogeneity. It also skirts the so-called structure–agency problem. What kind of *spontaneity* is possible in a social system that would normally require individual actors to respond to and modify the social pressures of the environment of which they are part? How is a spontaneous *order* possible if individuals are to transform their own relationships in contexts that are ontologically distinct from agency?

The above perspectives have added to our understanding of a range of issues concerning cluster evolution. For example, they have highlighted the potential adaptive utility of structural and behavioral diversity, the role of extra-cluster linkages, the dynamics of organizational foundings and failures, and the role of institutions as rule and convention setters, supporting selection and retention. By emphasizing the variability of cluster performance across time and space, they have contributed to the understanding of issues that are at the core of economic geography. And by showing – albeit often indirectly – that these issues involve social and cultural frames, they have pushed a view of cluster evolution as an essentially human enterprise, thus falling squarely within the purview of human geography. On the whole, however, researchers who have drawn on these perspectives have not been keenly interested in the *specifics* of human agency. By agency specifics I mean the actions individuals engage in and the frames they adopt when they make sense of the past, orient themselves to the future, and construct the present in light of the contingencies of the moment. Because the structural contexts of action and interpretation are themselves temporal (relational) fields, agency is 'analytically situated within the flow of time' (Emirbayer and Mische, 1998, p. 963) and thus inherently evolutionary. The challenge for evolutionary analysis is to explain the evolution of structures that enable individuals not only to act but to act *meaningfully*.

The most interesting aspects of human geography are probably those that are modified, prevented, enforced, or weakened by individuals who plan, improvise, reiterate, imitate, communicate, or forget. As agentic activities and orientations, these aspects should be at the core of an understanding of clusters as structurally and behaviorally complex socio-economic systems in which intentionality is everything and little is left to chance. However, many cluster researchers working from an evolutionary perspective have taken a rather distant view of agency. Complexity theorists emphasize the importance of diversity among elements of a complex adaptive system but they normally have little to say about how exactly the system self-organizes to enable 'symmetry-breaking', the construction of a 'spontaneous order', or the reproduction of a 'deep structure'. Population ecologists emphasize context in terms of the cultural and material properties of resource environments but are relatively insensitive to the micro-level processes and mechanisms by which environments are enacted and populations are maintained as meaningful social units. Co-evolutionary thinking tries to correct for this oversight by noting the fluidity of boundaries between levels of action, but it is relatively silent on exactly how cross-level connections are made and adjusted to, for example, prevent dissipative structures from turning into chaotic structures. What is needed for a more comprehensive evolutionary explanation is an understanding of how agents in the system interact with exogenous forces to ensure the survival of the emergent order, or to push for a new order.

I suggest that a better understanding of these issues calls for a more fine-grained and temporally sensitive analysis of human agency. The social-evolutionary perspective outlined below puts agency at the center of the analysis, using the general Darwinian principles of variation, selection, and retention to explain the direction and speed with which a complex entity, such as a cluster of individuals and organizations, evolves. The Darwinist algorithm[3] treats current structures as the outcome of an evolutionary process made possible by the competitive selection of heritable variations and the transmission of selected variations to future generations. The social-evolutionary perspective takes human intentionality seriously, without imposing essentialist assumptions concerning human nature. For evolution to work, one need not assume, for example, that humans are predisposed, genetically or otherwise, to engage in specific behaviors such as trust-based cooperation or self-interested competition.

3. Basic components of the social-evolutionary approach

Evolution is not a linear and expedient optimization process but a context-dependent and typically flawed process of adaptation to local contingencies, involving human agents who are intentional but limited in their capacities. The question of whether individuals are mindlessly or mindfully intentional is of interest to evolutionary theorists more at higher levels of analysis, where actors are seen as entities strategically engaged in collaboration, rivalry, identity building, and the like. Of the evolutionary perspectives

Table 10.1 Key concepts in the social-evolutionary framework

Concept	Description
Evolution	All components of clusters are subject to evolution, including the institutions, rules, and networks through which individuals operate. Social evolution is driven by the general Darwinian algorithm of variation, selection, and retention, analogous to biological evolution but with mechanisms that are distinctly social and vary across local settings.
Idea	Ideas are the unit of selection. They constitute the raw material with which individual actors interpret and construct their world. Evolutionary selection works on the phenotypic, behavioral expression of ideas.
Population	Joint dependence on the same resources (human attention) is the main defining criterion for idea populations. The distribution of ideas in a given population constitutes the ecological space within which ideas unfold.
Variation	Variation is the basis for selection and retention. Ideas change through mutation and recombination. Variations arise in a process that may be 'blind' with respect to the adaptive value of new combinations of ideas.
Human agency	Individuals are the translators of ideas. They are intentional but not always calculatively rational because of social biases, conformity pressures, and cognitive limitations. Agency is situated in a context of fundamental complexity and uncertainty, which it shapes in the process of adapting to it.
Mechanism	Mechanisms are the causal force connecting outcomes and initial conditions. Higher level mechanisms may be driven by mechanisms operating at lower levels. Mechanisms in the translation of ideas include imitation, improvisation, and discourse.

discussed above, some versions of institutional theory have perhaps moved the farthest towards a systematic consideration of deliberate agency, by studying the social construction of culture with reference to cognitive and political processes (DiMaggio, 1997). Also population ecologists have recently shown growing interest in agency, by, for example, exploring the social construction of identity as a strategic blueprint for organizational forms (Hannan et al., 2007). In the literature on clusters, deliberate agency is a topic for researchers studying clusters as 'epistemic communities' in which the actors organize themselves around the mastery of shared codes (Håkanson, 2005). However, when these researchers refer to tacit knowledge, experience, or novel business ideas, they normally stop short of examining the content of ideas and the accumulated knowledge and experience that ideas create. I suggest that new insights can be obtained from examining ideas in their own right, as well as the mechanisms that drive their evolution. Table 10.1 provides an overview of the key concepts in the idea-based social-evolutionary framework.

Ideas as the unit of analysis and evolutionary selection

That ideas matter in economic geography is probably beyond dispute, but it is rarely made explicit. Beliefs, assumptions, understandings, and uncertainty can make the difference between the success and failure of cluster policies. Ideas can promote cluster change or they can help maintain the status quo. We speak of good ideas when we marvel at the foresight of cluster planners, and we speak of novel ideas when we celebrate the imaginativeness and ingenuity of cluster entrepreneurs. Depending on the context, we have different interpretations of what constitutes good or novel ideas. Ideas are implicated in valuation and analysis. Some ideas are implicitly held, forming the taken-for-granted backdrop of interpretation. Other ideas are explicitly expressed, constituting the substance of communication and negotiation. Some ideas produce consequences in the background, others work in the foreground. If ideas are oriented to outcomes, they have cognitive content, such as ideas about the likely effectiveness of certain policy initiatives. Normative ideas consist of values and are typically not outcome oriented, such as notions about the distributional fairness of a particular governance system.

Some scholars have criticized theories of institutions and organizations for neglecting how ideas and meanings are implicated in change (Blyth, 1997; Ingram and Clay, 2000). The key point, they argue, is that what people believe (ideas) is just as important as what they want (interests), and what they want is often determined by what they believe. In some fields, such as political science and political sociology, there has been renewed interest in 'how ideas matter' (Campbell, 1998; see also the contributions in Goodin and Tilley, 2006), picking up where Talcott Parsons (1938) left off long ago. Economic geographers also refer to ideas as input factors in 'innovation systems', as the raw material of 'creative clusters', or as outcomes of strategic deliberations, but they have generally not problematized ideas as a subject of inquiry in their own right. The concern for ideas by cluster researchers who draw inspiration from population ecology, organizational evolutionary theory, and complexity theory has, at best, been latent. Ecologists studying population dynamics have generally looked at the outcomes of competition and cooperation, such as firm survival, but they have typically not studied directly the cognitive structure of a population of actors to answer questions about how the actors choose between different ideas. They have usually taken the existence of resource niches as given and have not asked how actors actually construct their world in these niches. When researchers in the field of organizational evolution invoke arguments from institutional theory, they generally refer to those that explain convergence and reproduction, while downplaying the older versions of institutional theory that take a more fluid perspective on struggles over resources and ideas (Hirsch and Lounsbury, 1997). Complexity theorists typically offer no direct account of the social constructions involved in the spontaneous process by which order is created. How do actors frame their actions vis-à-vis one another? What are the ingredients of the 'self-referencing framework' that guides decisions in a way consistent with the system's accumulated history? What exactly goes into actors' choices in constructing what are essentially contestable social structures? In other words, we need to know something about the substrate on which decisions, routines, competencies, and so forth depend. In the social and cultural domain of clusters, ideas would seem to qualify as such a substrate, with sufficient variation to make evolution possible and with sufficient longevity, fidelity, and fecundity to permit replication (Dawkins, 1976).

Ideas may be thought of as constituting the micro-foundation for the entities that contribute to cluster evolution. Being a foundational unit means that they are distinct from the entity of which they are part, such as rules and routines. The evolutionary question is whether or not an idea is sufficiently discrete to function as a unit of selection (see the debates in Aunger, 2000). If ideas are to function as replicators (Dawkins, 1976), in the sense that they have generative powers and can be transmitted in their entirety, they must be analytically divisible. Just as clusters, organizations, or routines are not established in a fully specified form and do not evolve *in toto*, individuals do not acquire information in an amalgamated form. But how do we put boundaries around an idea? One possible answer is that one can study ideas as discrete units, as long as their informational elements are sufficiently distinct to determine if the information acquired and passed on is novel, contains errors, has been changed, and so forth. This is what Burt (2004) did when he studied the likelihood that the ideas that information brokers have get noticed, are accepted as valuable, and are passed on. For this to happen, there must exist individuals with an attentive and interpreting mind. As long as the presence or absence of particular ideas is discernable in the minds of individuals, ideas can function as units of selection, making possible the description of clusters in terms of the cognitive elements underlying phonotypic representations, such as a common identity or a joint strategy. The only requirement for an evolutionary account of the transmission of ideas is that there are observable differences between them and that these differences can affect their reproductive success. The ambiguity surrounding the boundary (of ideas) question is the evolutionary condition that enables imagination, innovation, and new knowledge creation.

Studying ideas as the micro-foundation for entities driving the evolution of clusters has several implications. First, it serves to remind theorists that the analysis of cluster dynamics should go beyond a narrow focus on competition and cooperation. Selection operates along multiple dimensions, including criteria such as cognitive consistency, historical fit, or social legitimacy. Whether differences within a population of units at a given point in time have consequences for population evolution depends also on the malleability of the units. One may hypothesize about the robustness of ideas and the strength of the selection process. Background ideas, for example, are more stable than ideas operating in the foreground. Changes in the 'deep' cognitive structure of clusters, such as their core identity, are costly and potentially risky, as we know from clusters in East Germany that have struggled to reinvent themselves after German reunification (Heidenreich, 2005). A selection process that favors reproducibility works in favor of organizations and clusters that exhibit strong inertial tendencies. Selection criteria rewarding reliability and accountability might include ideas concerning, for example, the expectation that businesses invest in economically distressed regions. The long-term effect of such ideas may be to inhibit rather than strengthen the forces of change. A selection environment rewarding organizational reproducibility is one where innovation achieved through the transformation of existing patterns is relatively rare. Selection in this case will, over time, eliminate firms that alter their core features too often, independent of interfirm competition or cooperation.

Second, focusing on ideas draws attention to persistence and change independent of the *assumption* – evident, for example, in population ecology theory – of inertia. Studying cognitive micro-foundations highlights the different pace at which evolution proceeds at different levels of action. In general, evolution proceeds at a faster pace at lower levels of action, with outcomes that may differ from those at higher levels. For example, high failure rates at the level of firms do not necessarily jeopardize the survival of the cluster as a whole if, for example, strong institutional support systems remain in place (Staber, 1998). Or, the interpersonal linkages between specific innovators may be very unstable, without affecting the identity of the regional cluster as a whole (Cantner and Graf, 2006). The reason why evolution is generally faster at lower levels is that the diversity of units that evolutionary selection can explore is smaller for subunits than for whole entities. Thus, the probability of discovering the 'correct' configuration quickly is lower for higher level entities (Sober, 1984). This means that the fitness of ideas will evolve more quickly than the fitness of the cluster as a whole. Ideas themselves may vary in the speed of evolution. Some ideas, such as those related to local traditions, may evolve very slowly if errors in the replication process of traditions are rare (Shils, 1981). Others, such as those related to collective strategy, may evolve more quickly if contestable ideas are more prone to replication errors.

The third point concerns the relationship between the intended adaptation at the cognitive level of individual actors and the adaptive capacity of the cluster as a whole. For example, a policy of attracting resource-rich firms into the cluster may have the perverse effect of decreasing the fitness of the cluster if these firms, through their actions, destabilize the balance of power (Bathelt and Taylor, 2002). Competing ideas may help create a diverse repertoire of potential solutions, useful for innovation if they are enacted in ways that provide the system with freshness as opposed to complacency. A broad distribution of competing ideas may have a more invigorating effect on change than a narrow distribution of nearly identical ideas. The key adaptation challenge for actors in uncertain environments, such as those in new-media clusters, is not to imitate, codify, or routinize what successful firms are doing, but to create opportunities for redrawing and broadening cognitive categories by reconfiguring relational boundaries (Girard and Stark, 2002). The (in evolutionary terms) optimal cluster arrangement may be one that permits the co-existence of contradictory logics of interpreting and evaluating ideas and actions. System diversity, and the ambiguity that goes with it, remain a general condition for the growth of knowledge (Pavitt, 1998).

Variation

The concern for ideas as discrete units carries over into the discussion of variation, which is central to any explanation of evolution. Without variation, there can be no selection and retention and, therefore, no possibility of improving the fit between the organizational form of a cluster and its external environment. If variation is to enable evolution, it has to create new opportunities (there is more than one attempt to create change) and must be able to fail (some variations must be forgotten, given up, or lead to dead-ends) (Campbell, 1965). For example, new knowledge created in one cluster setting may diffuse – via labor mobility, firm relocation, and so on – to other settings where it gets mixed in with existing knowledge. The application of knowledge in diverse contexts will likely produce new information as to its performance implications. Some new ideas will be generated in this process, while others will be given up. The changing distribution of variable ideas throughout the evolutionary cycle is an issue of primary concern in discussions of epistemic and practice communities (Amin and Roberts, 2008).

Some researchers may prefer taking an essentialist approach to the question of ideational variation, by defining a small number of salient properties, such as ideas being instrumental or sentimental, social or economic. But this is problematic if no concrete guidelines are provided for distinguishing between trivial and central properties, and if it leads researchers to highlight average features and thus to overlook the entire range of (potential) variations. New variations of potential value may arise at the margins of a distribution, as would be expected, for example, in clusters containing firms with easy access to resources located in distant locations. Some investigators may take an agnostic approach to the question of ideational variation, in order to create sufficient analytical space for contemplating the entire range of potentially relevant variations. They might study, for example, the boundedness of ideas or populations of ideas in terms of the processes through which distinctions are maintained (Lamont and Molnár, 2002). To the extent that these boundaries are contested, the process of boundary construction is subject to evolution. For cluster researchers, the issue then would be to examine how actors tie boundaries together to create new distinctions around idea complexes underlying, for example, shared identity, forms of governance, or collective strategy.

Associated with the view that ideas evolve in an ecological space in which ideas compete for human attention (Heath et al., 2001) is the argument that populations of ideas exist as polythetic combinations of ideas that are sufficiently related to be recognizable as distinct ideational systems, such as those constituting a cluster's identity. This raises the question of how to determine which of the many ideas flowing throughout a cluster converge to a *dominant* cluster identity. The argument that clusters are distinguishable groupings of dominant ideas based on, for example, a core technology or a local tradition, is problematic if there is a risk of mistaking dominance for adaptive value. From an evolutionary perspective, what turns out to be dominant or valuable is difficult to predict, as individuals have a range of abilities they can use to accomplish a variety of goals in a system that, as a whole, may not have a single principle for variation and selective retention. Different clusters may be organized around different 'focal rules' (Schelling, 1960), used for different purposes, such as coordination, learning, or identity maintenance. Some of these rules may support dominant ideas that turn out to have no adaptive value.

Given that, at any point in time, clusters normally contain a mix of adaptive and maladaptive variations, the evolutionary question concerns the origins of new variations that might re-balance existing configurations. Network-theoretic arguments are typically invoked to theorize about the relational conditions that either enable or constrain the generation of new variations. Networks carry information and, depending on how they affect the perception, interpretation, and enactment of ideas, they help transform information into knowledge, while reconstituting the individual actor's mindset and the cluster's collective mindset in the process. It has become something of an axiom to argue that tightly woven social networks unduly constrain the remixing of routines (Staber, 2001), but the precise mechanisms involved in this process are not well understood. It is not clear, for example, whether change is enhanced by a thick network of strong ties connecting subsets of a cluster (bridging ties) or by a dense network of primary contacts (cohesive ties). The evolutionary outcomes of network structure are a function of the interaction between environmental conditions and the responses of cluster actors (Glückler, 2007) who are motivated by evolving ideas about strategy, identity, technology, cooperation, and so forth (Hite, 2005).

Under conditions of uncertainty, the new variations that emerge from this interaction are essentially 'blind' with respect to the fitness of the system (Campbell, 1965), an idea that is at odds with the (conventional) view of clusters as planned arrangements. Evolutionary theorists question not so much the existence of rational choice and planned interventions than the (often implicit) assumption that the consequences of actions are closely tied to intentions. The 'wisdom' of an idea – that is, whether it is adaptive or maladaptive – can be interpreted only retrospectively, once environmental conditions are known. Bradshaw's paradox, which states that we 'need to know the biological results before we can decide on the appropriate space to represent our compounds' (Nightingale, 2000, p. 337), applies to the entire innovation process in clusters. Human agency is important, through the purposeful projection of a vision of sorts and through the actions intended to fulfill it, but this does not imply predictability. Agency is an indeterminate process in which no actor has definitive knowledge and in which clear models are often lacking. People often make decisions based on cognitive distortions and perceptual errors (Levinthal and March, 1993). They select initiatives because they trust the information from small samples, they overestimate their analytical capabilities, and they interpret events ex post as something they responded to strategically when, in fact, they reacted haphazardly. This view of intentionality departs somewhat from that followed by complexity theorists who interpret the emergent properties of clusters as the *spontaneous* outcome of individual actions (Lindsay, 2005). It also challenges the view of evolutionary theorists who think of organizations as relatively stable bundles of routines. Procedural organizational memory may help to institutionalize certain ideas, protecting them from extraneous influences. But this does not imply that such ideas will (automatically) structure actual behavior.

While variation in the cognitive capabilities and imaginativeness of agents may be crucial to system evolution, existing institutional and social forces may reduce the diversity of outcomes from such variation. The task for evolutionary analysis is to assess the net balance between these opposing forces in particular instances. In dynamic systems such as clusters, it may be useful to consider blind variation in agency, together with chance events occurring in volatile environments, a baseline model with a null hypothesis and then to explain the emergence of new variations beyond stochastic developments. For this kind of analysis, all actions, interpretations, and decisions have *potential* relevance, and any attempt to create new knowledge, develop a new strategy, or build new relationships has *potential* value. The practical implication of this is that intervention should not aim at compensating for variations with unknown adaptive value, but should embrace these variations through, for example, supporting entrepreneurs located at the periphery of a cluster (Belussi et al., 2008), building institutions that encourage continuous experimentation (Grabher and Stark, 1997), or encouraging reflexive network management practices (Sydow, 2003).

Mechanisms of idea transmission

If the evolutionary approach is to deliver on its promise of explaining persistence and change, it must specify the causal mechanisms that operate at various levels in a given system. Theorists should not be content just with a process account of evolving socioeconomic systems. Understanding the causal mechanisms operative in such systems is essential, both in itself and for understanding process. Cluster researchers often use causal concepts when they refer to actors 'constituting' networks or 'enacting' beliefs that then 'diffuse' throughout the cluster, institutions 'enabling' or 'constraining' governance, or historical legacies 'embedding' firms in local social structures. In general, however, they are vague about the mechanisms by which beliefs are enacted, firms are embedded, or institutions constrain behavior. The problem of defining mechanisms arises because the various entities relevant to selection exist in a nested hierarchy. The evolution of an entity may be viewed in isolation (by treating it as a unit with self-contained mechanisms of replication), constitutively (by identifying the lower level mechanisms that drive its evolution), or contextually (by demonstrating how it fits into the structure of a higher level mechanism). Mechanisms also operate along a temporal dimension, with outcomes that may differ across levels. While some mechanisms, such as idea codification, have immediate and situation-specific effects, their aggregate and long-term effects (for example, in form of accumulated experience) may vary considerably, depending on initial conditions and the constellation of all other mechanisms at work in specific empirical instances.

A causal mechanism includes the particular feature of the unit of selection that brings about outcomes, as well as the processes connecting outcomes and initial conditions (Hedström, 2005). In evolutionary explanations of organic systems, the competition among genes and the contribution of competition to species fitness is seen as the key causal mechanism. In the context of socio-economic entities like business clusters, the repertoire of mechanisms is potentially large and diverse. Social evolutionary theorists have approached the question of mechanisms in contrasting ways. On one side are those who follow a strong version of biological analogy, assigning 'selfish', gene-like qualities to ideas in the sense that ideas have an 'interest' in and can affect their own propagation (Dawkins, 1976). For these theorists, it is the competition among ideas and their contribution to social fitness that is considered the central causal mechanism. This implies that it is no longer individual actors, in their role as decision-makers or communicators, who should be the primary research focus, but the properties of ideas that enhance their own survival chances in an environment containing competing ideas. Research has investigated how 'ideas acquire people' (Lynch, 1996) in form of myths, legends, ghost stories, and the like (Heath et al., 2001). This line of research holds promise for those interested in explaining how ostensibly harmful ideas – for example, notions like secrecy and distrust (Staber, 2007) – can persist over the long term. This research does not deny the role of individual agents in proposing, modifying, and communicating ideas. It rather draws attention to the possibility that some ideas may spread even though individuals are unaware of, indifferent to, or even opposed to them (Dennett, 1995). Whether the study of ideas as replicators can mature into a full-blown science of culture remains unclear at this point (Aunger, 2000).

Other theorists have adopted a view that is more appreciative of the deliberate intervention of human agents who provide an interpreting mind and thus give meaning to the information contained in ideas. They propose that any novel idea that an individual accepts must be 'translated' (Czarniawska and Joerges, 1996) – perceived or modified in some way – to be more consistent with other ideas in a given context. Virtually any idea can become stable within a social group if it is sufficiently supported by social norms, which themselves are composed of ideas (Richerson and Boyd, 2005). A variety of translation mechanisms may be involved, such as imitation, improvisation, and discourse. Individuals may imitate others as closely as possible if, for example, they believe it is in their best political interest to do so (motivated by social conformity needs). Or, they may imitate others because they perceive so much uncertainty that they don't know what else to do but to follow those in their surroundings (motivated by uncertainty reduction needs). When ideas travel from site to site by imitation, the recipients may enact them in different ways, thus contributing new variation. Translation may also involve deliberate improvisation when individuals detect minimal information, discover unintended consequences of their actions, and correct their mistakes until they find a reasonably acceptable modus operandi. Translation in this case is motivated by a search for plausibility rather than accuracy. This ensures that the necessary adjustments can take place, while preserving the central structure of the social system (Tilley, 2005). Actors may also use existing language and texts (in form of written documents, visual displays, speech acts, and the like) when engaging in discourse with others. Discourse is a mechanism for structuring not only interaction but also perception through the use of linguistic rules, symbols, and analogies (Gibson, 2005). Through discourse, actors draw on, transform, and reproduce their social reality. For example, the academic discourse about clusters has produced a wide range of ideas concerning flexible specialization, cooperation, collective learning, and the like. These ideas are in a constant struggle for survival, as they compete for human attention and are retained and passed on in form of, for example, best practice and policy initiatives.

The concern for translation mechanisms reflects a version of agency that is more socially engaged, with more discretionary, evaluative, and transformative possibilities than what is offered in many evolutionary accounts at more aggregate levels. While social transmission is different from genetic transmission, because of the capacity of the mind to reinterpret and reapply the information passed on, information processing behavior is also subject to the possibility of mutation. The relative probability of further replication in a changing environment affects the speed and direction in which the relevant population evolves. Idea-based information varies, is selected, and is copied, through a variety of mechanisms and with more or less imperfection, and this is all that is needed for evolution to occur.

In research practice, it may be very difficult to determine which particular mechanism is operating in a given empirical instance. The analysis of mechanisms is complicated by the fact that they exist in a nested hierarchy. Higher level mechanisms may be driven by mechanisms operating at lower levels. For example, the argument that knowledge circulates in a cluster through the interfirm mobility of employees (Agrawal et al., 2006) needs further specification because mobility may be motivated by a variety of cognitive and social logics. To what extent causal regression to lower level mechanisms is analytically required, is a question of debate not only in the philosophy of science (Craver, 2001). It is also a concern for cluster policy makers who grapple with the question of the level of action at which intervention is most likely to be successful.

The study of mechanisms forces attention to the actual processes connecting causes and effects. This improves theoretical understanding, for example by not mistaking spurious associations for real causal relationships, which is a danger especially when the goal is to explain macro-level phenomena. A systematic concern for mechanisms also helps avoid the proliferation of seemingly disparate theoretical concepts if the same mechanisms work in different cluster processes. The problem is that in socio-economic systems as complex as clusters several mechanisms will probably operate simultaneously, and they may counteract one another in their effects on particular outcomes. Although it is difficult to disentangle multiple mechanisms empirically, cluster theorists ought to find the notion of mechanism appealing and worth pursuing in their research.

4. Concluding remarks

The impermanence and fluidity of regional clusters calls for theoretical explanations that appreciate the various ways in which human agency is embedded in those aspects of clusters that make them 'fluid spaces' (Mol and Law, 1994). Employing Darwinian principles at a high level of abstraction, of which the evolutionary processes in the biological and social realm are but specific instances (Dennett, 1995), opens up opportunities for exploring clusters as idiosyncratic social evolutionary systems in which macro-level institutions (MacKinnon et al., 2009) are as important as micro-level processes related to, for example, politics (Bathelt and Taylor, 2002) or learning (Staber, 2009). The emerging view in research on clusters is that variability is a central, and often, useful feature, produced by a dynamic mix of competition and cooperation, and framed by rules and understandings that are continually contested and updated. Evolutionary theory, grounded in culture rather than biology, can refine this view by adding improved explanations of the processes of change, without neglecting patterns. Evolutionary explanations shift focus from what a cluster is to how it is accomplished. As we move to a notion of clusters as an ongoing accomplishment we need a notion of agency to match. In this chapter, I suggested that an idea-based social constructionist approach offers significant explanatory leverage, in a way that cross-fertilizes different variants of evolutionary thinking with a systematic concern for the details of agentic orientation and action. The challenge for researchers is to explore how ideational alternatives are constituted, how the actors choose from a pool of ideas, and how the distribution of ideas in a bounded system changes over time and across settings. As a general proposition, this would direct researchers to expect a wider range of contingencies in agency. Cluster theorists would be less concerned with deviation from a particular norm and more fascinated by variation.

To be useful for an understanding of cluster evolution, any variant of evolutionary theory must achieve at least three objectives. First, it must explain the dynamics of how cluster firms and structures change over time. In doing so, it must recognize the underlying variability of the components making up firms and populations. In particular, it must appreciate human agency, leaving room for cognitive diversity and discretionary possibilities, and recognizing that human actors have both stable and fluid properties. Second, it must show a systematic concern for the mechanisms that generate both persistence and change, rather than merely describe statistical regularities. It must permit an analysis of mechanisms with sufficient generality to be able to advance causal generalizations about recurrent processes at all levels of action. And third, it must consider the ever-present possibility that clusters evolve in ways that do not improve system performance or do not serve the interests of all their members equally well. The explicit recognition of dysfunctional developments helps to avoid the functionalist reasoning evident in much of the cluster literature. If one defines clusters by their purported functions, such as innovation or knowledge sharing, there is the obvious need to explain the nonappearance of functions as well. One would also need to suggest functional alternatives, as well as the mechanisms by which evolutionary processes produce different outcomes, both adaptive and non-adaptive.

The focus on different facets of agency, such as cognitive capacity and practices, helps to protect against functionalist reasoning by showing that actors not only have various kinds of ideas, often inconsistent and fleeting, but also routinely make mistakes in interpreting them and passing them on. Imperfections in idea transmission may not only create new variation in the social system, but may also alter the intentions and understandings of individual actors in their ongoing dialogue with the situations unfolding around them. Just because an idea is enacted on a sustained basis does not mean that it must be successful in resolving the problem that may have triggered it. People are subject to social biases, they engage in struggles over control, and they strike political compromises. In their concern for institutions, researchers should study not only processes that reproduce routines successfully but also those that cause routines to be abandoned or forgotten. These processes include also 'failed' ideas that, by opening up new spaces, contribute to the differential reproduction of behaviors across time and space.

While evolutionary theory cannot predict *exactly* how actions and patterns will evolve, it *can* make predictions that guide future research. The task for evolutionary theorists is to explain how the variation, selection, and retention of ideas – as well as other relevant units – have led to the patterns that we recognize as clusters. The analysis of change is then a matter of testing hypotheses that link the content of ideas to those features of the agents and their environment which increase or decrease the likelihood that they will be repeated and that they will displace those previously selected. This requires data sets with sufficient detail to capture relevant variations in actors, ideas, practices, and social mechanisms, across levels of aggregation and in all relevant domains (Holmes and McKelvey, 2005). It also requires dynamic data structures that enable the separation of random noise from true underlying time trends and regularities. Of obvious interest are historical data, including those collected through qualitative methods, in order to explain how things currently are, what social actors think they are doing, and what it means to them. Research on clusters would benefit greatly from temporally sensitive data-gathering and analysis methods, as well as more process-oriented qualitative methods, such as process-analytic narratives (Griffin, 1993) and process-tracing methods (Bennett and Elman, 2006). To be sure, the data requirements of such methods are so severe that many researchers may resort to simulation studies instead (Chang and Harrington, 2005).

At a meta-theoretical level, the social-evolutionary approach proposed here draws on a variety of perspectives employed in cluster research, such as interpretivism, resource dependence, and learning theory. Different perspectives shed light on unique aspects of variation, selection, and retention in a way that no single approach can do. When combined, different perspectives reveal the complexity, multiplexity, and probabilistic nature of the forces that shape the life of clusters. By contemplating multiple rationalities of actors and by examining action at multiple levels, different perspectives provide useful raw material for the further development of the evolutionary approach. The objective of an evolutionary research program, one grounded in society rather than biology, is not to replace other perspectives but to offer a platform from which the insights of other perspectives can be interpreted, in order to improve our understanding of how regional clusters came to be the way they are.

Notes

* I am grateful to Ron Boschma and an anonymous reviewer for their critical comments on a first draft of this chapter. Any remaining errors are mine.

1. I refer to clusters here as a geographically concentrated set of related populations of organizations with sufficient unit character to be recognizable as a more or less integrated 'community', having what Wittgenstein (1958, p. 32) called 'family resemblance'. Depending on theoretical focus or policy interest, such communities are variously referred to as regional learning systems, industrial districts, innovative milieux, or hot spots.
2. Adaptations are any traits that increase an entity's chances of survival, as opposed to advantages that improve its general 'well-being'. Many cluster studies focus on advantages, such as interfirm collaboration, institutional support, or shared identity, which may or may not have adaptive value for the entity.
3. An algorithm is a set of step-by-step instructions, like the building rules that govern the activities of

wasps. The Darwinian explanation of evolution relies on mindless algorithms driving bottom-up processes (Dennett, 1995).

References

Agrawal, A. I. Cockburn and J. McHale (2006), 'Gone but not forgotten: knowledge flows, labor mobility, and enduring social relationships', *Journal of Economic Geography*, **6**, 571–91.
Amin, A. and J. Roberts (eds) (2008), *Community, Economic Creativity, and Organization*, Oxford: Oxford University Press.
Aunger, R. (2000), *Darwinizing Culture: The Status of Memetics as a Science*, Oxford: Oxford University Press.
Bathelt, H. and M. Taylor (2002), 'Clusters, power and place: inequality and local growth in time–space', *Geografiska Annaler*, **Series B/84**, 93–109.
Belussi F., S. Sedita and L. Pilotti (2008), 'Learning at the boundaries for industrial districts between exploitation of local resources and exploration of global knowledge flows', in R. Leoncini and S. Montresor (eds), *Dynamic Capabilities between Firm Organization and Local Systems of Production*, London: Routledge, pp. 181–215.
Bennett, A. and C. Elman (2006), 'Complex causal relations and case study methods: the examples of path dependence', *Political Analysis*, **14**, 250–67.
Blyth, M. (1997), 'Any more bright ideas? The ideational turn in comparative political economy', *Comparative Politics*, **29**, 229–50.
Boschma, R. and K. Frenken (2006), 'Why is economic geography not an evolutionary science? Towards an evolutionary economic geography', *Journal of Economic Geography*, **6**, 273–302.
Boschma, R. and R. Martin (2007), 'Editorial: Constructing an evolutionary economic geography', *Journal of Economic Geography*, **7**, 537–49.
Burt, R. (2004), 'Structural holes and good ideas', *American Journal of Sociology*, **110**, 349–99.
Campbell, D. (1965), 'Variation and selective retention in socio-cultural evolution', in H. Barringer, B. Blanksten and R. Mack (eds), *Social Change in Developing Areas: A Reinterpretation of Evolutionary Theory*, Cambridge, MA: Schenkman, pp. 19–48.
Campbell, J. (1998), 'Institutional analysis and the role of ideas in political economy', *Theory and Society*, **27**, 377–409.
Cantner, U. and H. Graf (2006), 'The network of innovators in Jena: an application of social network analysis', *Research Policy*, **35**, 463–80.
Chang, M. and J. Harrington (2005), 'Discovery and diffusion of knowledge in an endogenous social network', *American Journal of Sociology*, **110**, 937–76.
Craver, C. (2001), 'Role functions, mechanisms, and hierarchy', *Philosophy of Science*, **68**, 53–74.
Czarniawska, B. and B. Joerges (1996), 'Travel of ideas', in B. Czarniawska and G. Sevón (eds), *Translating Organizational Change*, New York: Aldine de Gruyter, pp. 13–48.
Dawkins, R. (1976), *The Selfish Gene*, Oxford: Oxford University Press.
Dennett, D. (1995), *Darwin's Dangerous Idea*, New York: Simon & Schuster.
DiMaggio, P. (1997), 'Culture and cognition', *Annual Review of Sociology*, **23**, 263–87.
DiTommaso, M., P. Paci and S. Schweitzer (2004), 'Clustering of intangibles', in P. Bianchi and S. Labory (eds), *The Economic Importance of Intangible Assets*, Aldershot: Ashgate, pp. 73–102.
Emirbayer, M. and A. Mische (1998), 'What is agency?', *American Journal of Sociology*, **103**, 962–1023.
Gibson, D. (2005), 'Taking turns and talking ties: networks and conversational interaction', *American Journal of Sociology*, **110**, 1561–97.
Girard, M. and D. Stark (2002), 'Distributing intelligence and organizing diversity in new-media projects', *Environment and Planning A*, **34**, 1927–49.
Glückler, J. (2007), 'Economic geography and the evolution of networks', *Journal of Economic Geography*, **7**, 619–34.
Goodin, R. and C. Tilley (eds) (2006), *The Oxford Handbook of Contextual Political Analysis*, Oxford: Oxford University Press.
Grabher, G. (2009), 'Yet another turn? The evolutionary project in economic geography', *Economic Geography*, **85**, 119–27.
Grabher, G. and D. Stark (1997), 'Organizing diversity: evolutionary theory, network analysis and postsocialism', *Regional Studies*, **31**, 533–44.
Griffin, L. (1993), 'Narrative, event-structure analysis, and causal interpretation in historical sociology', *American Journal of Sociology*, **98**, 1094–133.
Håkanson, L. (2005), 'Epistemic communities and cluster dynamics: on the role of knowledge in industrial districts', *Industry and Innovation*, **12**, 433–63.
Hannan, M., L. Pólos and G. Carroll (2007), *Logics of Organization Theory: Audiences, Codes, and Ecologies*, Princeton, NJ: Princeton University Press.

Heath, C., C. Bell and E. Sternberg (2001), 'Emotional selection in memes: the case of urban legends', *Journal of Personality and Social Psychology*, **81**, 1028–41.
Hedström, P. (2005), *Dissecting the Social*, Cambridge: Cambridge University Press.
Heidenreich, M. (2005), 'The renewal of regional capabilities: experimental regionalism in Germany', *Research Policy*, **34**, 739–57.
Hirsch, P. and M. Lounsbury (1997), 'Ending the family quarrel: toward a reconciliation of "old" and "new" institutionalism', *American Behavioral Scientist*, **40**, 406–18.
Hite, J. (2005), 'Evolutionary process and paths of relationally embedded network ties in emerging entrepreneurial firms', *Entrepreneurship Theory and Practice*, **29**, 113–44.
Hodgson, G. (2007), 'Institutions and individuals: interaction and evolution', *Organization Studies*, **28**, 95–116.
Hodgson, G. and T. Knudsen (2006), 'Why we need a generalized Darwinism, and why generalized Darwinism is not enough', *Journal of Economic Behavior and Organization*, **61**, 1–19.
Holmen, M. and M. McKelvey (2005), 'Systematic evolutionary studies of regional restructuration', *European Urban and Regional Studies*, **12**, 209–28.
Ingram, P. and K. Clay (2000), 'The choice-within-constraints new institutionalism and implications for sociology', *Annual Review of Sociology*, **26**, 525–46.
Lamont, M. and V. Molnár (2002), 'The study of boundaries in the social sciences', *Annual Review of Sociology*, **28**, 167–95.
Levinthal, D. and J. March (1993), 'The myopia of learning', *Strategic Management Journal*, **14**, 95–112.
Lindsay, V. (2005), 'The development of international industry clusters: a complexity theory approach', *Journal of International Entrepreneurship*, **3**, 71–97.
Lorenzen, M. (2002), 'Ties, trust and trade: elements of a theory of coordination in industrial clusters', *International Studies in Management and Organization*, **31**, 14–34.
Lynch, S. (1996), *Thought Contagion: How Belief Spreads Through Society*, New York: Basic Books.
McKelvey, B. (1997), 'Quasi-natural organization science', *Organization Science*, **8**, 352–80.
MacKinnon, D., A. Cumbers, A. Pike, K. Birch and R. McMaster (2009), 'Evolution in economic geography: institutions, political economy and adaptation', *Economic Geography*, **85**, 129–50.
Malmberg, A. and P. Maskell (2002), 'The elusive concept of localization economies: towards a knowledge-based theory of spatial clustering', *Environment and Planning A*, **34**, 429–49.
March, J. (1991), 'Exploration and exploitation in organizational learning', *Organization Science*, **2**, 71–87.
Martin, R. and P. Sunley (2007), 'Complexity thinking and evolutionary economic geography', *Journal of Economic Geography*, **7**, 573–601.
Mol, A. and J. Law (1994), 'Regions, networks, and fluids: anaemia and social topology', *Social Studies of Science*, **24**, 641–71.
Nelson, R. and S. Winter (1982), *An Evolutionary Theory of Economic Change*, Cambridge, MA: Belknap Press.
Nightingale, P. (2000), 'Economies of scale in experimentation: knowledge and technology in pharmaceutical R&D', *Industrial and Corporate Change*, **9**, 315–59.
Parsons, T. (1938), 'The role of ideas in social action', *American Sociological Review*, **3**, 652–64.
Pavitt, K. (1998), 'Technologies, products and organization in the innovating firm: what Adam Smith tells us and Joseph Schumpeter doesn't', *Industrial and Corporate Change*, **7**, 433–52.
Richerson, P. and R. Boyd (2005), *Not by Genes Alone*, Chicago: University of Chicago Press.
Rigby, D. and J. Essletzbichler (1997), 'Evolution, process variety, and regional trajectories of technological change in U.S. manufacturing', *Economic Geography*, **73**, 269–84.
Romanelli, E. and O. Khessina (2005), 'Regional industrial identity: cluster configurations and economic development', *Organization Science*, **16**, 344–58.
Schelling, T. (1960), *The Strategy of Conflict*, Cambridge, MA: Harvard University Press.
Scott, W. (2001), *Institutions and Organizations*, 2nd edn, Newbury Park, CA: Sage.
Shils, E. (1981), *Tradition*, London and Boston: Faber and Faber.
Sober, E. (1984), *The Nature of Selection*, Chicago: University of Chicago Press.
Staber, U. (1998), 'Interfirm cooperation and competition in industrial districts', *Organization Studies*, **19**, 701–24.
Staber, U. (2001), 'The structure of networks in industrial districts', *International Journal of Urban and Regional Research*, **25**, 537–52.
Staber, U. (2007), 'A matter of distrust: explaining the persistence of dysfunctional beliefs in regional clusters', *Growth and Change*, **38**, 341–63.
Staber, U. (2009), 'Collective learning in clusters: mechanisms and biases', *Entrepreneurship and Regional Development*, **21**, 553–73.
Sunley, P. (2008), 'Relational economic geography: a partial understanding or a new paradigm?', *Economic Geography*, **84**, 1–26.
Sydow, J. (2003), 'Management von Netzwerkorganisationen: Zum Stand der Forschung' [Management of

network organisations: the state of research], in J. Sydow (ed.), *Management von Netzwerkorganisationen* [Management of Network Organisations], Wiesbaden: Gabler, pp. 293–354.
Sydow, J. and U. Staber (2002), 'The institutional embeddedness of project networks: the case of content production in German television', *Regional Studies*, **36**, 223–35.
Tallman, S., M. Jenkins, N. Henry and S. Pinch (2004), 'Knowledge, clusters, and competitive advantage', *Academy of Management Review*, **29**, 258–71.
Tilley, C. (2005), *Identities, Boundaries, and Social Ties*, Boulder, CO: Paradigm Publishers.
Wittgenstein, L. (1958), *Philosophical Investigations*, 3rd edn, Upper Saddle River, NJ: Prentice Hall.

11 Evolutionary economic geography: regional systems of innovation and high-tech clusters

Philip Cooke and Carla de Laurentis

1. Introduction

Evolutionary economic geography integrates numerous strands of heterodox spatial fields. Some parts of even 'new neoclassical' theory, such as increasing returns, aspects of endogeneity in growth processes, and notions of abiding spatial disequilibrium as the resultant of increasing returns to spatial scale (Krugman, 1995), provide interesting insights into why some regions are rich and others poor, to echo Malthus's famous question of Ricardo (Landes, 1998). But this is all, for there are few other distinctive contributions to be made from that quarter. One sub-field of regional science that is more squarely compatible with an evolutionary approach is that dealing with regional innovation systems (Braczyk et al., 1998). It is avowedly neo-Schumpeterian, translating that author's resolutely aspatial economist's mode of analysis and bringing further life to it. Indeed, elsewhere it has been argued that one reason why it continually attracts fascination is that regional innovation systems analysis gave a boost to more general innovation systems thinking. This, according to Carlsson (2007), is the dominant innovation studies field since refereed articles began appearing in the mid-1990s (Cooke, 1992).

The reason for this is explored in section 1 of the chapter, arguing that the conceptual perspective of industrial economics, often referred to as 'industrial dynamics' in the innovation studies literature, is vertical – down the sector, as it were, from the vertically integrated large firm to its suppliers and support organisations. In the old days of the 'Industrial Age' when multinationals evolved as vertically integrated behemoths, pursuing industrial organisational processes that could be captured in simple 'S' shaped curves by the likes of Vernon (1966), this fitted a linear theory of innovation. But, as is well known, that model began breaking down not long after it was conceptualised, and large firms began to outsource even essential requirements, chiefly to respond to Japanese 'lean production', although the formulation of that academically came some twenty-plus years after Vernon (1966) in the 'lean production' bible of Jones et al. (1990). The geographical imagination is more horizontal, finding little difficulty even seeing much of what hitherto passed for vertical integration as far from complete. One only had to have spent time in 1950s to 1960s Birmingham or Coventry, as did one of this chapter's authors, rurally raised, on frequent visits to relatives who had migrated from the declining Welsh coalfields to seek a better life, to experience the shock of seeing numerous (usually black-painted) brand-new truck chassis with unenclosed engine and wheels, and hung-on temporary licence plates, being driven through the city-centre to one of many distant city coach-building specialists to be finished, to experience 'agglomeration' at first-hand (Boschma and Wenting, 2007).

This chapter then moves into an analysis of regional evolution, especially regarding systemic innovation. Important concepts are juxtaposed involving the more interesting fringes of new neoclassical economics and a resolutely systems analytic approach to urban and regional studies, whose origins also lie in the 1960s, based in general systems theory, as in much natural science, engineering and technology, evolving from there as a complex analytical and guidance model expressed in the discourse of 'systems theories of planning' (Chadwick, 1971; McLoughlin, 1969). This was the integrated 'substance and process' approach to spatial analysis that modernised a professional urban design approach that had hardly changed since the Pre-Raphaelites, by virtue of modelling urban and regional systems mathematically, utilising gravity models, and the like, to predict behaviour. It was over-ambitious, complicit with what are now perceived as destructive biases in favour of, for example, suburbanisation, separation of land-uses, primacy of vehicular mobility in cities, and the destruction of traditional heritage and ambient city environments. However its underlying conceptual analysis, unlike its prognostications, was effective at facilitating representations of complex process realities in non-trivial ways.

The same can be said of the more mainstream innovation systems approach, which populates its perspective on the object of interest with core systems concepts such as 'networks, nodes and interactions' including feedback and 'institutional learning' (Lundvall and Johnson, 1994). Accordingly, as noted, it is open to small elements (increasing returns, asymmetric information, principal–agent relations, and possibly transaction costs) of new neoclassical economics or spatial econometrics. This is also because of *its* neo-Schumpeterian interest in variety, search, selection, routines, trust, and embeddedness. But it goes further in sharing interest with both regional innovation systems and, for example, Italian industrial district theory, in collaboration, innovation, learning, path dependence, institutional change, disequilibrium and knowledge intermediation practices of institutions and organisations, including firms (Boschma and Frenken, 2006). These 'bring life back into economic(s) geography' to paraphrase Hodgson (1993). In what follows, the chapter reprises and updates the case for an evolutionary approach over any alternative approach to the understanding of spatial process. Finally, having made the connection to the evolutionary study of regional innovation, one strand of the broadening field of evolutionary economic geography, the chapter finishes by devoting empirical attention to search and selection procedures conducted by high-tech businesses in UK ICT agglomerations of which the M4 motorway corridor is one, and East Anglia, centred on Cambridge, is the other. The interest here lies in the discovery of practices of possibly adverse search and selection. That is, what is normally presumed to be an asset of agglomerative behaviour – indeed it is said to be a main driver of the 'cluster craze' (Asheim et al., 2006) – namely access to prized 'knowledge spillovers' and 'swift trust' (Sabel, 1995), turns out to be associated with quite high transaction costs and extortionate rents, the apotheosis, in other words, of diseconomies of agglomeration.

2. Theorising evolutionary economic geography

Neoclassical theory hypothesises that economic actors are homogeneous, rational, non-opportunistic and capable of calculating best value optimal decisions in a world without uncertainty. The firm is the centrepiece of this ratiocination. Here, economic actors transform inputs into outputs represented mathematically as a production function given appropriate technology and, through the market, an external price mechanism for estimating costs and calculating optimal profit. Technology is exogenous, but also endogenous for new neoclassicals, firms being assumed to have comparable technological competences (Romer, 1990). Firms' other important characteristics are similar, remaining so during their life cycle. In a world without uncertainty, contracts can be complete and fully definable. For the neoclassicals, competition is pure and perfect, no barriers exist to market entry, and firms have equal access to resources, knowledge and information that are freely available, and optimal labour and capital input coordination is affordable.

In essence, the neoclassical model is one in which the firm's decisions and activities are driven by the price mechanism in a world of pure and perfect competition. Fundamental questions, which neoclassical theory provides no answer to, concern growth and development, coordination and knowledge. In the neoclassical tradition, firm growth is a sign of imperfect competition, one of the many paradoxes arising from a prejudice in favour of equilibrium as the natural economic state. Neoclassical theory also fails to explain mechanisms observed in the real world such as partnership, networking, or oligopoly because it assumes the uncertainty that these practices denote is non-existent. Similarly, heterodox proposals regarding the normality of firm and individual choice as satisficing rather than optimal decision-making cannot satisfactorily be allowed without relaxing the absence of uncertainty principle, as the new neoclassicals in fact commonly do (Simon, 1962). Nor can the impact of phenomena such as history, routines, location of research and production centres or advantage of technological and individual skills superiority be modelled. Despite recognition that much technological advance is endogenous (Romer, 1990), the 'new' neoclassical models are still 'mechanical': uncertainty, conflicts among experts, unexpected results that mark the innovation process are not integrated (Nelson, 1995). Moreover the endogenous outcome remains the technological artefact, and classical form of 'congealed capital' rather than the more embedded and embodied form of innovation.

As suggested in the introduction to this chapter, evolutionary economics has a different way of conceiving economic actors, firms and markets. It places a pronounced emphasis on history, routines, and interactions and influences of environment and institutions. Firms are conceived of as specific actors, or preferably, agents (Pavitt, 1984). Moreover, following Simon's (1972) observation of the impossibility of any single agent having omniscience with regard to information appropriate to optimal decision-making, given the complexity of their environment and conflicts of interest that disallow profit maximisation, agents are not assumed to be able to compute optimal solutions and even less to predict other agents' behaviours because of uncertainty (Alchian, 1950; Heiner, 1983; Knight, 1921). Hence, bounded rationality (Simon, 1962, 1972), satisficing behaviour and differing expectations (Hahn, 1952 [1984]; Rosenberg, 1982) are taken to be normally expected practice. In evolutionary economics, firms are not uniform but distinctive, and utilise differentiated capabilities, one of which is knowledge, another is administration and management. As Penrose (1959 [1995]) saw, both knowledge and organisation play significant roles. Increasingly, even compared with the era of Penrose (1959 [1995]), knowledge and externalised knowledge networks were seen to have risen in importance because of the increase in *innovation* as a factor of the firm's self-constructed competitive advantage. The organisation of a firm as an efficient and effective form of administration managed to optimise this, and its other resource capabilities had never been a stronger factor in its development, aided increasingly by institutional and organisational learning. From an evolutionary perspective, firms learn from their own experience but also from other firms and organisations they interact with and with whom they exchange knowledge both organisational and technological. Evolutionists recognise firms have histories, path dependencies and development trajectories. They may show capabilities of survival and prosperity in maintaining a relatively unchanging market location, as was the experience of many banks, for example, or they may have a special capability of transforming themselves to fit new market locations by virtue of their foresight competence, like DuPont or Nokia, as cases in point. The latter explore new paths of growth to exploit; the former adapt more to new demands by incorporating learned routines, outsource technological requirements, learn of new opportunities, and adapt to new constraints. If not, they succumb to competition and become an acquisition, possibly of some new, more innovative vehicle (e.g. private equity business), and if they cannot evolve to meet new market exigencies they exit the market.

As this chapter focuses empirically on some evolutionary learning characteristics of clusters, it is worthwhile postulating some possibly original dimensions differentiating a more neoclassical from an evolutionary line of reasoning and hypothesising between the two. Here we refer to the different perspectives on the role of knowledge spillovers in the emergence of clusters. Keep in mind clusters are more than agglomerations of sectoral neighbours in geographic proximity. Connectivity through communication, trust, reputation, favour-exchange and other forms of collaboration is involved, as shown in the empirical section of this chapter. Knowledge spillovers are part of the adhesive in these arenas of high social capital but also competition. The neoclassical approach to knowledge spillovers and clusters as drivers of growth is represented in the Marshall-Arrow-Romer position that privileges *specialisation* of knowledge and expertise as the growth driver. Hence the fewer inter-sectoral knowledge spillovers that reduce the effectiveness of absorptive capacity the better. Single clusters in a possibly random 'darts in a dartboard'-like space would be consistent here.

An evolutionary approach would be more akin to Jacobs (1969) and her well-known proposition that diversity is the dynamic driving innovation. From a cluster perspective, this leads to a hypothesis about cluster mutation as an emergent property of 'Jacobian clusters'. The evolutionary terminology here would be that of related variety and proximity effects hastening lateral absorptive capacity among mature and embryonic clusters. Precisely this phenomenon is found in places like California, North Jutland, Wales (Cooke, 2008a) and, as Boschma (2005) shows, in Emilia-Romagna. In the former cases convergence then cluster emergence show specific path dependence (see Chapter 3 in this book) from ICT through biotechnology to clean technology clusters in California, agricultural and marine engineering through wind turbines and solar thermal energy clusters in North Jutland, and wireless telephony through medical technology to biotechnology. Predecessor clusters emerged after Schumpeter's fifth innovation category of 'railroadisation' had opened up the respective territories in the 1800s. In Wales agro-food stimulates emergence of renewable energy cluster emergence (bioethanol and biomass) while electronics, automotives and aerospace share common photonics spaces, skills and competences. Cluster-related variety based on engineering skills typifies 'third Italy' clustering according to Boschma (2005). Clearly this is a powerful explanation for regional evolution and associated policy thinking.

3. Evolutionary theory and regional evolution

Having introduced the distinctiveness of Jacobian clusters, this section takes that insight as a basis for examining first the regional, then the local cluster forms of regional evolution. As indicated, evidence has been accumulated to show that regions can improve their prosperity by adjusting or even transitioning in relation to their historic path dependence and history. This is by no means easy, but possibly as a result of the over-ambitious prospectus of 1960s systems planning and reflection on the need to prioritise system 'levers', if they experiment by moving away from uniform policy prescriptions and evolving an integrated substance-process approach to regional development, with constant monitoring and adaptation, regions can become innovative systems. They may previously have been non-innovative systems, 'locked in' to an apparently evaporating industrial paradigm, or they may have been fragmented industrial regions possessed of diverse and unconnected economic elements, and for these the evolutionary challenge to search for and select a survival strategy on which to build a success strategy is at its most acute. The temporal dimension, in the absence of accomplished governance of this process, may be long enough to be economically fatal and the region never develops, but it may be foreshortened by judicious application of the knowledge and organisational competences Penrose (1959 [1995]) sees as characteristic of the organisationally and cognitively sophisticated firm.

In evolutionary theory, it will be recalled, firms are conceived of as collective organisations with a variable degree of internally generated and externally learned resource-development capability. Regions and regional development have more in common with this perspective than with the neoclassical world of homogeneous, atomistic units of rational utility maximisation. Accordingly, the evolutionary theory of the firm and the region enjoy greater conceptual complementarity. Conceiving them as differentiated, making use of variable proportions of knowledge and capability inputs and benefiting from methodologies both for learning and, more importantly, knowledge-generation, on which economic advantage may be constructed, are key in this. Such capabilities are path-dependent (Arthur, 1994) but not predetermined, they can be learned, thus widening the range of feasible innovation opportunities affecting economic progress. Path dependence is criticised as inclining towards the deterministic, but it can be shown to make a significant contribution to understanding cluster evolution through related variety in 'Jacobian clusters' (see section 2 above and section 6 below). Related variety can be seen as rather static and even conservative in its reliance on official statistics, but it can be dynamised in cluster theory by the evolutionary notion of 'cluster mutation'. Networks are shown from the research reported in the penultimate section of this chapter to be the determining feature of clusters defined as geographically proximate social capital for purposes of firm evolution. More generally, such relations are also globalised through 'distant networks' that, among other things, help explain 'open innovation', that is, global outsourcing, even of research if not yet of innovation where spatial proximity still seems vital.

Unlike the neoclassical world of isolated utility-maximisers for whom technology and learning gains are still largely exogenous, purchased off the shelf, the evolutionary world is one in which innovative, imitative, unpredictable and Pasteur's 'fortune favours the prepared mind' effects occur. Illustratively, but empirically true, a region that might be deemed, in effect, redundant since its population had largely left to seek opportunity elsewhere, its agriculture was uncompetitive and there were in any case low linkages with existing manufacturing industry, itself non-innovative, might turn around even in these least auspicious conditions. In a world of 'peak oil' and energy insecurity, the empty fields of the dying agricultural economy might be reinvigorated by the judicious interactive innovation of selecting them to be suppliers of bioenergy to the manufacturers who hitherto were un-innovative and thought they had nothing to discuss with the representatives of the remaining farmers. Regions like the American mid-west, former East Germany, not to mention central-southern Brazil, are experiencing such an evolutionary regional development process because what was once a marginal and far-fetched, systemic idea – a post-fossil fuel economy – has moved closer to the mainstream.

These changes have the consequence that, through interaction with other firms and agencies, the economic environment is itself modified as well as exerting its own modifying effects. The diffusion of both codified and tacit knowledge among firms in relatively equal relationship to one another, especially where they are competitive outside their domestic base but complementary, or even collaborative within it, is an important source of constructed advantage for small firms. But constructed it has to be; for example, the aforementioned linking of traditional agro-food production with the world of large-scale energy production requires construction of new interactive networks. An interesting question is how precisely this happens. How, in other words, may related industrial variety arise from unrelated? Yesterday, farmers and automotive workers may have had little reason to interact closely except through the anonymity of the market where food and tractors were purchased. But today, in areas that were recently thought unviable in agro-food terms, like the region of Mecklenberg-Pomerania in Germany, parts of northern England, and elsewhere, interactions among representative associations and research institutes from agriculture, energy and automotives have begun discussions and announced investments in biofuels (Goodall, 2007; Jürgens et al., 2007). The evolution of this 'revealed related variety' is why, for example, Italian industrial districts have proven capable of maintaining a competitive edge in traditional sectors despite competition from low wage, less developed economies. They have *systemic* process elements and knowledge flows inherent within and between them. In periods of relative economic stability, the system generates and absorbs externalities of the kind neoclassical theory assumes to be efficiently internalised in the institution of the firm (Boschma, 2005).

Such relationships are not *hierarchical*, they are *heterarchical*. Heterarchy is the condition in which network relationships pertain, based on trust, reputation, custom, reciprocity, reliability, openness to learning and an inclusive and empowering rather than an exclusive and disempowering disposition (Cooke, 2002). However, heterarchy does not operate in a vacuum. Modern regional development theory, even more than evolutionary economic theory, emphasises the importance of the socio-cultural milieu (Maillat, 1995) within which network forms of inter-firm organisation are embedded (Granovetter, 1985). We are not here talking about community in a simplistic and generic way, rather about routine practices and mentalities of entrepreneurship in the context of a commercial community. As Marshall (1919) put it:

> . . . good work is rightly appreciated, inventions and improvements in machinery, in processes and the general organisation of the businesses have their merits promptly discussed; if one person starts a new idea, it is taken up by others and combined with suggestions of their own; and thus it becomes the source of further new ideas. (Marshall, 1919, p. 271)

Theorisation of system weaknesses in heterarchic, localised forms of economic coordination has gone furthest in industrial district theory. Two key system weaknesses have recently been identified in the canonical neo-Marshallian form of the industrial district. Writing from an evolutionary economics perspective, Varaldo and Ferrucci (1996) have identified the following developmental blockages presently visible in the district form of development. First, because of strategic cooperation between firms in districts there develops, of necessity, a common set of strategic expectations about behaviour among mutually dependent entrepreneurs. Institutional memory, rules, routines, ways in which mutual expectations are regulated, relationships ordered and rules governed – these come to converge so that dissonance within the district becomes muted. This, the absence of dissenting voices, reproduces district culture – in the commercial sense – but may delay *strategic* creativity at critical points when a rapid response to the need for innovation is required. If we remember that innovation is defined as the *commercialisation* of original knowledge, as distinct from invention, which is the original knowledge itself, then the need for rapid response becomes obvious.

Second, this is not a problem until the district system experiences an exogenous shock, such as a stabilisation, or even, as occurred in the global recession of the early 1990s, a contraction in demand. This can cause a number of panic responses: competition may become cut-throat and destructive; low prices make investment in restructuring impossible; reduced demand limits the high flexibility levels associated with district firms, revealing hidden costs as margins are cut and more standardised production is resorted to; firms seek cheaper offshore suppliers and even production locations; retail customers are more able to negotiate favourable contracts for themselves; new technologies may represent a threat where they are incongruent with the technical know-how of district entrepreneurs. All these reactions can be experienced in non-district settings too, and there is even limited evidence that in Italy, during the 1990s recession, district firms fared better than similar ones outside districts (Brusco et al., 1996).

However, these forms of industrial organisation have proven remarkably resilient, adapting skills and technologies and responding to inauspicious external conditions. Most recently, in the mid-2000s some were thriving by absorbing large numbers of Chinese workers and entrepreneurs, at least 30,000 in Prato, one of the most traditional textile towns in Italy (Dei Ottati, 2009). Founded in its modern form as an initially impoverished local economy based on the recycling of wool, in effect from rags, to be re-woven into fabric by multitudes of small artisans, the district has dealt with external acquisitions of some key firms, the innovation imperatives of new software-controlled machines, and much competition from low-wage economies in core markets. The arrival of many Chinese entrepreneurs has brought new skills and upgrading of the commodities on offer, including designer clothing and accessories. There is some irony in that a major migration flow is from China's clothing and textiles 'industrial villages' in southern China to which the Italian Communist Party, which then ruled in Tuscany (and Emilia-Romagna), often sent consultant advisers on how to build industrial districts in the People's Republic (Becattini, 2001).

The systemic dimension of innovation

Most economic development involving enterprise support for small and medium enterprises occurs in situations far removed from the industrial district model of development.

Varaldo and Ferrucci (1996) concluded that the future of Italian district firms depended more and more on their capacity to make links to non-district firms displaying strategic competitive advantage, including, crucially, networks of global firms. Instead of tightly defined districts, their model for the future was of more loosely defined 'clusters' of inter-firm relationships. These were thought capable of taking advantage of the incremental innovation made possible in periods of relative stability where close networks and what we now call 'related variety' could facilitate rapid information diffusion and learning. But for the more dynamic, strategic innovations by means of which global competitiveness is sustained, firms in local networks need to be in touch, not necessarily directly, but through the supply-chain, with global networks. Clearly no one saw the likelihood of globalisation producing an involuted evolution of certain districts (especially clothing and textiles-related) with Asian immigrant upgrading to the fore. Nor at that time was the establishment of new districts in low-wage countries like Romania envisaged. Finally, as Becattini and Coltorti (2006) make clear, Italy's industrial districts in the post-war epoch have generally out-performed the rest of the economy in terms of firm and employment growth.

If we broaden the discussion of economic development from the local to the regional level, keeping the possibility of an exploration of the systemic relationships open, then we are drawn into reflection on the notion of 'cluster' since it is the systemic rather than simply agglomerative nature of the phenomenon that is of potential interest and the cluster concept rests on that characteristic. We may briefly explore two related but distinctive elaborations of the basic cluster idea and, in the process, say more about the *systems of innovation* approach to thinking about how these may be coordinated. The *systems* dimension arises from the 'membership' of the network comprising the cluster. Minimally, the *innovative region*, especially the high-technology kind, may be expected to have agglomerations of new technology businesses that have the characteristics, denoted above, of clusters. It is crucial not to see 'cluster' and 'regional innovation system' as synonymous. This is for three reasons: in scale terms clusters are seldom regional but local; clusters have very different governance mechanisms from regions, ranging from associative 'clubs' to little more than regular market interactions; and regions may contain many clusters (or none) as well as other organisational forms of industry. Clusters too vary, and not only in terms of governance. In Bottazzi et al. (2002) and Paniccia (2006) useful, but very firm-focused typologies range from 'horizontally diversified' through 'Smithian' supply chain to 'oligopolistic' and 'science-driven'. Conceivably, examples of each might be found in a region. The key point about the concept of *cluster* is that while it will consist of firms, large and small, in a lead industry and supporting activities, it is the interactive nature of such firms, ranging from doing business to doing favours, that is its generally distinctive feature.

Further network relationships may exist proximately or virtually, with research and higher education institutions, private R&D laboratories, technology transfer agencies, chambers of commerce, business associations, vocational training organisations, relevant government agencies and appropriate government departments. This constitutes the basis for an integrative governance arrangement. The club, forum, working party, consortium or partnership model is what typifies this *associative* (Casson, 1995; Cooke and Morgan, 1998; Hirst, 1994) approach towards enhancing the commercial community. From such arrangements, institutional learning and innovation gains may more readily be acquired. So the second dimension, foreshadowed above, is the *associative governance* of the cluster. Conceptually, this involves a major shift from state-regulation of economic affairs to a degree of self-regulation by responsible groups in economy and society, but not strictly 'liberal market' governance. In Hirst's (1994) formulation this means ceding some aspects of economic governance to associations at large capable of managing certain aspects of communal provision (such as vocational training or technology transfer), supported by appropriate financial mechanisms. It also implies decentralised, transparent and consultative governance. Institutional learning is a crucial part of an associative approach. It presumes no fount and origin of all wisdom; rather it assumes the processes of economic development and especially innovation are interactive ones in which institutions on the user side (e.g. customers) may be as important as producers (e.g. scientists) of the innovations in question. Localised cluster evolution can be significantly assisted by associative institutions and organisations such as these. They in turn may occasionally act collectively where a regional agency has capabilities to supply further, more strategic, support. In the next section the chapter moves into a report of research findings into the question of whether firms in clusters perform better than their ICT sectoral equivalents if they exist in clusters and collaborate on research and innovation activities.

4. Measuring the effect of proximity and collaboration on firm performance

The research to be reported on here administered postal questionnaires to structured samples of UK healthcare biotechnology (not covered in this chapter; for methodological detail of the project, see Cooke et al., 2007) and ICT companies (264 hardware, software and services respondents) enquiring about comparative firm performance of collaborator and non-collaborator firms in and outside clusters. Clusters were defined as being located among sectoral neighbours and actively collaborating with them in a general, possibly informal way or cooperating specifically and contractually on some topic such as R&D, knowledge transfer or marketing, as well as networking distantly in the same respects. Non-collaborators defined themselves by reporting they did no such partnering, formal or informal, with firms or organisations inside or outside the cluster. They claimed they only engaged in market exchanges with customers or suppliers. This addresses a matter of key practical importance to this contribution, namely whether firm performance is affected by business 'clustering'.

Given what has already been written above regarding the *specialisation* emphasis in the work of neoclassicals and that perspective's neglect of non-market exchange interactions, it would be unlikely to hypothesise collaboration though it might hypothesise contractual cooperation. The richer evolutionary approach hypothesises both, and as it turns out, most fruitfully (on related issues, see Boschma and Frenken, 2006 for further discussion). In fact, from the evolutionary economic geography perspective, the really pressing question is whether it has a better explanation for clustering than the neoclassical approach, and if so why. To tackle this, the data first report more general interactive characteristics of firms in clusters. It is important to establish the extent to which firms consider themselves 'collaborative', whether in geographical proximity or not. In particular it is interesting and important to separate collaborator performance from general performance, and key indicator data for ICT are presented in Table 11.1, which compares key performance indicators for firms that collaborate or do not collaborate

Table 11.1 Collaboration and performance of UK ICT firms – key indicators

Selected performance indicators	All respondents	Collab-orators	Non-collaborators	Collab-orators		Non-collaborators	
				Cluster	Non-cluster	Cluster	Non-cluster
Mean employment	105	180	40	57	53	19	41
Employment increase 2000–03	36%	40%	34%	45%	41%	39%	39%
New products/ services 2000–03	80%	88%	73%	89%	87%	74%	69%
Patents 2000–03 +/−	0%	2%	−2%	6%	4%	−3%	0%
Turnover increase 2000–03	61%	69%	55%	70%	76%	49%	70%
Mean R&D expenditure/ turnover	16%	17%	14%	19%	13%	10%	11%
R&D expenditure increase 2000–03	31%	32%	30%	39%	16%	22%	33%

Notes:

1. Respondents to the questions on collaboration/non-collaboration were fewer and had a smaller mean than all collaborators. Columns 5–8 have N = 55, 40, 71 and 44 respectively.
2. The few large multi-plant firms in the sample frequently reported collaboration with affiliates or sister plants, which clearly fell outside the project's defined interest in inter-organisational and inter-firm collaboration and cooperation. Hence they are excluded from Table 11.1.

Source: CASS ICT Collective Learning Survey.

and their cluster/non-cluster location. Analysis was undertaken with respect to measures of firm performance, namely employment change, turnover change, research and development expenditure change, and innovation, all between 2000 and 2003. In this respect, levels of innovation were measured by asking firms about the number of new products/ service and changes to products/services in the past three years from survey time, the number of patents announced, R&D activities and the firms' capacity to introduce new products/services compared to competitors.

Table 11.1 shows that in all economic performance, collaborating UK ICT firms' mean performance is generally better than the mean scores in the respondent group as a whole, consisting of both collaborators and non-collaborators. Thus collaborators have superior performance regarding market share, capacity to introduce new products and services, higher R&D as a share of turnover in 2003, more employees per firm and greater turnover, showing a higher share of firms that recorded an increase in both

Table 11.2 Collaboration and performance of UK ICT firms – correlations

	Collaboration effect on performance	Respondent perception of collaboration	Perceived innovation effect
Capacity to introduce new products/ services	0.059	0.213**	0.024
Market share improvement	0.273**	0.181	0.112
Patent announcements	0.158*	0.100	0.075
Turnover increase	0.172*	0.236*	0.199*
Employment change	0.013	0.182*	0.145
R&D change	0.038	0.107	0.086
Improvements of company's best products	0.102	−0.096	0.323**
New products for company but not new for market	0.029	0.117	0.142
New products new for company and new for market	0.230**	0.152*	0.203**
Number of new products/services	0.102	0.011	0.157*
Number of changes in products/ services	0.081	−0.003	0.096

**Correlation is significant at the .01 level (2-tailed); *correlation is significant at the .05 level (2-tailed).

Source: CASS ICT Collective Learning Survey.

figures between 2000 and 2003. Thus collaboration clearly pays in most dimensions of measurable firm performance. Hence we continue to believe on the basis of our evidence that collaboration provides a competitive advantage. Of course, high performance may attract collaboration, which would be an excellent evolutionary inference. Accordingly, it is not unreasonable to propose that ICT firms engaging in collaborative activity with others are more capable on the R&D and patenting input side of the innovation relation and they benefit on the output side with greater market share. This is also confirmed in Table 11.2 which shows how collaboration is significantly and positively associated with key performance and innovation indicators.

A further key question is the extent to which ICT collaborators – not forgetting non-collaborators – are found consciously locating in clusters. The answer provided below shows a picture where geographical (cluster) proximity for UK ICT firms is important, as shown by the number of non-collaborating firms that consciously decide to co-locate in a cluster (56 per cent), but whether this is proven to be beneficial for firms' performance is somewhat unclear. The operation of knowledge spillovers seems to be important, where substantial numbers of non-collaborators are found in clusters indicating that there is a 'knowledge spillovers' attraction effect even for those who envisage non-collaborative relations with their neighbouring firms. These may be assumed to be those seeking to exploit knowledge that is 'in the air'. Interestingly, collaborators in clusters perform only marginally better, but consistently so, than those not in clusters, except on R&D expenditure increase 2000–03 where collaborators increased most. This could be a temporary peak, or it could mean they experience a cluster effect for R&D expenditure if R&D expenditure increase levels remain high over time. Yet a further interpretation is that they are inefficient spenders of R&D investment that is only associated with marginally better overall performance than is the case for non-collaborators.

Focusing on the collaborators' side, Table 11.1 shows that for some indicators of economic performance, collaborators in clusters perform better than collaborators in non-clusters. Collaborators in clusters tend to have superior performance regarding higher R&D as a share of turnover in 2003 (19 per cent compared to just 13 per cent), a higher number of firms recording an increase in R&D expenditure between 2000 and 2003, more firms announcing patents, and a greater number of patents announced in both 2000 and 2003. Clustered collaborators tend to be bigger than non-clustered collaborators and have a higher number of firms that increased their employment size between 2000 and 2003. However, while mean turnover of the two sub-groups is similar (£5 million in 2003), 76 per cent of non-clustered collaborators increased their turnover between 2000 and 2003. It can be concluded that collaborators in clusters tend to be superior on the R&D and patenting input side. However, with regard to the output side the higher investments in inputs do not benefit firm performance.

Turning attention to the non-collaborators, it is perhaps surprising that a significant number of non-collaborators consciously locate in clusters. However, the data reveal that non-collaborators in clusters perform better than their counterparts in non-clusters in just some innovation indicators (employment increase and new products/services). In contrast, non-collaborators in non-clusters spend more on R&D, and there is a higher proportion of firms that increased their R&D expenditure and turnover growth in the 2000–03 period. It can be argued that clustering can provide competitive advantage to non-collaborators as the non-collaborators that co-locate in geographical proximity are smaller in size (an average of 19 employees compared to 41 for non-clustered firms and £2 million average turnover compared to £5 million for the isolated non-collaborators). This seems to be a confirmation that economic spillovers are available as even neoclassical literature predicts in cluster settings. However it must be questioned whether notions like 'collaborators' and 'non-collaborators' would enter their econometric radar. Nevertheless it would be consistent with their 'knowledge spillovers' attraction effect since small firms would be more rational in utility-maximisation terms to seek out such spillovers than more self-contained larger ones. In this respect, diseconomies of scale effects might be discounted. However, as shown in Table 11.1 and, later, Table 11.2, this may help innovation but does not necessarily benefit firms' performance.

Table 11.2 presents correlations measuring associations between performance variables and collaboration effects and perceptions. Thus, in a novel way we compare expected outcomes of collaboration with realised firm performance on these variables. This helps us get at motivations for collaboration. The results show the following: first, collaborators display relatively high and statistically significantly improved market share, wholly new product or service (to firm and, crucially, market) innovation, and to a lesser extent, turnover and patenting improvements (consistent with Table 11.1). Interestingly, performance indicator expectations from collaboration coincided somewhat with the actuality, measured on the variable 'capacity to introduce new products or services', but actual performance on that variable was lower and not statistically significant. This may be interpreted as a kind of 'over-optimism' variable where reality produced less from

Table 11.3 Proximity, cognitive and innovation advantages for UK ICT firms

Proximity indicator	All respondents, %	Collaborators in cluster, %	Non-collaborators in cluster, %
Swifter, clearer knowledge exchange	79	83	72
Reduce interaction cost	70	66	75
Facilitates informal communication	87	89	83
Reduce uncertainty	59	68	36
Facilitating collective learning	48	55	36
Innovation cooperation in cluster	23	26	20

Source: CASS ICT Collective Learning Survey.

collaboration than expected even though there were respectable improvements in wholly new product innovation. A similar conclusion can be made regarding turnover increase. It occurred but not as much as expected. Conversely, wholly novel innovation occurred more than expected, as did market share and patenting improvements but not employment. On innovation specifically, perceived and actual performance were marginally out of line, with wholly novel innovations being realised more in fact than was expected to be the case by firms.

Table 11.3 concerns the value of proximity and answers questions regarding 'all respondent' and 'collaborator–non-collaborator' 'cluster–non-cluster' performance. The first five rows are firm responses as to what their respondent (CEO or R&D/innovation manager) has experienced, beneficial or not, from cluster-based collaboration. It compares collaborators and non-collaborators in clusters on the same indicators. Row six reports, as the acid test, how much innovation cooperation, defined as contract-based and legally binding partnering, actually occurred. Collaborators favour spatial proximity to a greater extent than the respondent group as a whole and that of non-collaborators. Respondents answering these questions were low in number, with even those emphasising clustering cooperation for innovation a minority compared to those conducting such activities with intra-firm cooperation (larger firms), and intra- or even extra-UK innovation cooperation. Spatial proximity is thought beneficial by clustered firms that collaborate as it facilitates informal communication (89 per cent), facilitates knowledge exchange (83 per cent), and reduces uncertainty (68 per cent) and interaction costs (66 per cent).

Interestingly, the reasons that motivate proximate location differ among collaborators and non-collaborators in clusters. Most non-collaborating firms preferred clustering to reduce interaction costs, including transaction costs, whereas most collaborators valued this less. Non-collaborators also consider reducing uncertainty and facilitation of collective learning as less important than collaborators do. But on speeding up the knowledge exchange process and facilitating informal communication there is little difference between the two groups, as shown in Table 11.3. The main anomaly requiring explanation here refers to the possibly strong neoclassical interpretation above that

Table 11.1 indicates non-collaborators enter clusters to access knowledge spillovers. To begin with Table 11.3 shows some strong support for swifter knowledge exchange, but also shows less importance, as noted, placed on the core neoclassical presumption of such entry 'reducing uncertainty'. Finally, it is noteworthy, if unsurprising, that despite all the enthusiasm and practical action taken by collaborators and non-collaborators to access communicative connectivity in a cluster, far fewer actually get into contractual innovation cooperations.

Puzzlingly, 20 per cent of avowed non-collaborators also engage in contractual cooperation. From interviews conducted subsequent to the analysis of results with a representative sub-sample of the respondent population, two explanations arise. First, they perceive arm's length contractual relations as 'cooperation' and second, they actually do get into contractual innovation and other forms of cooperations as defined in this research to some extent, since they are truly 'opportunistic'. Cognitively dissonant or rationally utility-maximising? Perhaps firm-species mutation through learning to search for and select opportunity from the cluster ecosystem would be the most plausible explanation. But it also has to be entertained that as active non-collaborators they find themselves more excluded than they expected given their main expressed knowledge spillovers interest in recruiting talent or contractual opportunities. Alternatively what is clearly 'in the air' in the cluster is information about patenting and R&D that is not of as much interest to them or is beyond their absorptive capacity. Probably a combination of both lies close to the heart of the explanation, perhaps a 'diseconomies of scale' firm-cluster problem for non-collaborators is being experienced compared to the reverse for collaborators (also Table 11.1)

Hence we find convincing evidence from these results relative to the following three key dimensions. First, firms that collaborate perform better on nearly all performance indicators than firms that do not. Collaboration thus gives to firms in these industries an added competitive advantage. Second, collaborating firms in clusters perform better than collaborators not in clusters. Thus collaboration is good for business but geographical proximity is best. This means the cluster begins to take on the characteristics of what we wish to call 'constructed advantage'. This is a dynamically derived form of advantage constructed on the static qualities of agglomeration, which is transformed into a cluster by interactivity. Given these firms are in an ICT platform consisting of computing and communication hardware, software and services businesses, they derive evolutionary 'energy' from *related variety*. Finally, the cluster offers an unexpectedly large portion of even non-collaborating (56 per cent) ICT firms' constructed advantage. This arises from their conscious aspiration to access knowledge spillovers from the interaction effects and knowledge 'free-riding' opportunities available to firms within earshot of other incumbents with whom they have no intention of collaborating. The possibility that recruitment of talent is an element of knowledge spillover advantages being sought by such firms must also be taken into account. However, Table 11.3 shows they are frequently disappointed and some evolve to fit the cluster ecosystem by becoming collaborators against their avowed intent. Recall the significantly smaller employment size of these firms, which suggests that though this may not have been their primary interest, needs must. In general, the constructed advantage of the knowledgeable cluster thus derives from its local linkages and conveys degrees of competitive advantage directly and indirectly to its collaborators and non-collaborators alike. This is underlined by data showing non-collaborators in clusters are small but perform better than non-collaborators of any size outside clusters.

5. Contrasts in stylised interpretations of non-collaborator clustering: neoclassical and evolutionary

We have identified a somewhat unexpected practice by a significant portion of small and medium-sized enterprises in UK ICT. This is that there is sufficient attractiveness to locate in the midst of what are known to be clusters containing specialised and related variety firms with whom inward locators do not seek to collaborate in any intentional way. For the privilege, they are probably paying up to three times the land and labour rents they would pay in a not-too-distant science park environment outside the main clusters (e.g. M4 corridor, Cambridge or Oxford). They are mostly small, even micro-firms rather than even medium-sized ones. Their only expressed reasons for this locational practice are possibly to hear of sub-contracts or of skilled labour availability. To repeat, this is not to 'piggy-back' a possible consortium engaged in innovative actions, of which there are numerous ones in such settings, such as *Symbian* for 'Bluetooth' and 3.5/4.0 generation mobile telephony in Cambridge (Cooke and Huggins, 2003). Rather they are superficially 'free-riders' hoping to benefit from moderate 'knowledge spillovers' for which they are willing to pay up to a 300 per cent locational premium in conditions of great uncertainty. The uncertainty comes from the strong likelihood that to receive such spillovers they would have to have some knowledge with which to trade, which on the face of it seems unlikely, or worse, that they might expect knowledge in a cluster to be 'free-flowing' when in reality it may be more likely to be preserved in the 'club' atmosphere of locations such as Cambridge, where some such incumbents (e.g. in computer games) were socially excluded from networks involving the rather exclusive Cambridge Network Ltd, a firm established in 1986 on the San Diego CONNECT model to enhance network-based knowledge-flow among incumbent members (Cooke and Huggins, 2003).

We may stylise neoclassical and evolutionary interpretations of such, by no means exceptional, firm-practice, according to criteria listed in Table 11.4. First, with regard to motivation; second, with regard to uncertainty; third, relating to utility; fourth, information (resources); fifth, expectations (e.g. of cluster); sixth, price; and finally, the verdict as to whether such practice might be deemed economically rational or irrational – for

Table 11.4 Stylised neoclassical and evolutionary interpretation of non-collaborative clustering

Criterion	Neoclassical	Evolutionary
Motivation	Externalities	Knowledge spillovers
Uncertainty	High	Learning
Utility	Profit optimisation	Knowledge search
Information	Low	Selective
Expectations	Specialisation	Variety
Location price	Very high	Spillover cost
Rationality	Irrational	Rational

example, the practices of non-collaborators getting low information (Table 11.3) compared to their expectations, as well as not placing a particularly high value on 'reducing uncertainty'. Some neoclassicals might see this as irrational, possibly as 'adverse selection' and inconsistent with neoclassical rational utility maximising or optimising norms.

6. Concluding remarks

This chapter has proposed empirical evidence that gives confidence about the following key observations. First, and in theoretical terms, there is an advance in our understanding of the persistence and conceivable reinforcement of an asymmetric economic geography of prosperity and accomplishment. In an evolving and intensifying knowledge economy, science-driven and otherwise technologically sophisticated economic activity gives rise to demands on industry organisation that reinforce collaborative activity among smaller knowledge-intensive businesses, on the one hand, and between smaller, smart firms and university laboratories towards customer (and supplier) firms, many of which can, on the other hand, be shown to be large or even transnational corporations. This is important and original support for the thesis that *regional knowledge capabilities* increasingly determine the distribution of growth regions, currently favouring those that gain increasing returns from asymmetric knowledge distribution that assists in the construction of regional advantage in terms of talent recruitment and retention, spatial knowledge quasi-monopolies, and 'R&D outsourcing' or 'open innovation.' In UK ICT and biotechnology such features are pronounced, with key bioregional capabilities attracting these advantages to clusters like Cambridge and Oxford, while for ICT, London with its satellites in the M25 and M4 corridors is the dominant market-led magnet.

We also found the evolutionary perspective far superior in explanatory power in apparently non-utilitarian circumstances that nevertheless make sense when analysed from the evolutionary point of view. This is promising from a policy as well as a more academic viewpoint since it seems policy-making from a neoclassical point of view produces often counter-intuitive interpretations. This is, of course, because of the poverty of neoclassical perspectives both on knowledge and innovation analyses. This is perhaps surprising given that numerous founding fathers of neoclassical economics have emphasised the importance of both, but owing to the unnatural restrictiveness of modelling in that field neither has seriously been explored by its adherents. These are sub-disciplinary fields where evolutionary economic geography is more or less theoretically, conceptually and empirically unchallenged.

Thereafter, underlining the previous point, this research has tended to find support for the superiority of collaboration in respect of a variety of performance indicators, and clustered cooperation for innovation being supported more by the collaborating part of the firm sample than the respondent group as a whole. This broadly applies in ICT and biotechnology, but as we have seen, less regarding clustering for innovation activity by ICT than biotechnology firms, and much more for research interactions by biotechnology than ICT firms. Research in ICT is less of a cluster-driver than innovation activity, but the latter is not as pronounced as supply-chain innovation stretching global and intra-firm interactions, the latter partly a function of differing firm-size between the two samples. The one thing that appears to be almost transparent, especially in the ICT data, is the superiority for firm performance of collaborative knowledge exchange and innovation activity over stand-alone competition, even for the large, dominating firms in biotechnology, though possibly somewhat less for ICT firms, a few of which made it into the UK ICT respondent group.

Finally, there is a broad research agenda arising from this chapter in relation to the evolutionary economic geography of clusters, but space only permits the elaboration of three sub-fields. The first and still unanswered issue concerns 'cluster emergence'. In other words how can it be convincingly demonstrated that some embryonic spatial agglomeration has reached the point when it can be said to be a cluster? There is obvious leverage from solving this problem from the academic and, particularly, the policy perspectives. Is 'critical mass' an appropriate notion, what does it mean, and how can it be measured in ways that have wide applicability? Could evolutionary game theory be a means of probabilistically modelling cluster emergence? Finally, given clusters are defined in terms of their networks, distinguishing them from agglomerations that lack such dynamic spillovers, is cluster emergence best understood by application of evolutionary network theory (e.g. Cantner and Graf, 2006)? A second sub-field of great relevance to the fuller understanding of asymmetries in regional evolution centres on the idea that has been tested empirically to a limited degree in four case-settings, and concerns spatial variability in the presence of Jacobian clusters, and for that matter MAR clusters, and finally no clusters. Could the cluster mutation evolutionary process be a key to regional prosperity arising from proximate related variety? In the studied cases it seems to be so. In North Jutland particularly, it has a recorded beginning with Jutland's regional innovation through 'railroadisation' (Kristensen, 1992; Schumpeter, 1975) in the nineteenth century. The path dependencies identified through cluster mutation are as follows:

- Clean technology is path dependent on agricultural and marine engineering (e.g. wind turbine blades replicate plough and propeller design) near Aarhus and Aalborg.
- Biotechnology (BIOMEDICO) is path dependent on wireless ICT (NorCOM) and medical technology at Aalborg University.
- Wireless technology is path dependent on traditional ship-to-shore marine technology.
- Agro-food became established with the 'railroadisation' of Jutland; organic agro-food is a reaction against conventional intensive food production in Jutland (mostly pig and dairy).
- Furniture is path dependent on 'railroadisation', craft schools (350) and the local forestry tradition.
- Fashion clothing evolved for women from craft schools providing skills for farmers' wives in textiles.
- Modern fish equipment and pipework engineering is path dependent on traditional fishing and marine engineering centred on Aalborg.

A version of cooperative entrepreneurship with high social capital characterised the enterprise model of the Jutland pioneers and this has evolved into modern forms of collective entrepreneurship to the present day. Finally, what are the key policy mechanisms that usefully assist cluster emergence and evolution? The evaluation of eight VINNOVA Vinnväxt Programme cluster-building projects in Sweden suggests some common themes. Among the most important and frequently observable of these are: a pre-existing emergent agglomerative phenomenon; associative governance, leadership and finance; growth markets; and 'ahead of the curve' research and innovation knowledge (Cooke, 2008b). These mini-innovation system 'clusters' range from agro-food, through steel, robotics, industrial controls, robotics and fibre optics to healthcare services and biotechnology. Hence this is a robust test, conducted after the third year of cluster existence, but it is unclear whether all will survive and whether there is Swedish exceptionalism or maybe a new 'Swedish paradox' if all or any do. Hence there is a need for evaluative research, possibly utilising reference class forecasting (Flyvbjerg, 2008) to assist. This uses evolutionary evaluative modelling to establish actual costs and successes or failures of projects rather than traditional ex ante cost-accounting models that habitually overrun predicted costs and performance estimates.

Acknowledgements

We would like to thank the organisers, Ron Boschma and Ron Martin, and participants at the first Evolutionary Economic Geography workshop at St Catharine's College, Cambridge, UK in April 2006 where this chapter was first presented. In particular, our discussant, the late Bent Dalum, is thanked for his astute advice to 'edit down'. We also thank the reviewers for advice to 'edit down further' but also 'edit up' here and there. Finally the first author thanks the second for her great assistance in conducting the survey administration and analysis on the ESRC funded research project Ref: RES-000-23-0192 'Collective Learning in the Knowledge Economy: Milieu or Market?' The resources of ESRC are also gratefully acknowledged.

References

Alchian, A. (1950), 'Uncertainty, evolution and economic theory', *Journal of Political Economy*, **58**, 211–22.

Arthur, W.B. (1994), *Increasing Returns and Path Dependency in the Economy*, Ann Arbor: Michigan University Press.

Asheim, B., P. Cooke and R. Martin (eds) (2006), *Clusters and Regional Development*, London: Routledge.

Becattini, G. (2001), *The Caterpillar and the Butterfly*, Florence: Felice de Monnier.

Becattini, G. and F. Coltorti (2006), 'Areas of large enterprise and industrial districts in the development of post-war Italy: a preliminary survey', *European Planning Studies*, **14**, 1105–38.

Boschma, R. (2005), 'Proximity and innovation: a critical assessment', *Regional Studies*, **39**, 61–74.

Boschma, R. and K. Frenken (2006), 'Why is economic geography not an evolutionary science?', *Journal of Economic Geography*, **6**, 273–302.

Boschma, R. and R. Wenting (2007), 'The spatial evolution of the British automobile industry: does location matter?', *Industrial and Corporate Change*, **16**, 213–38.

Bottazzi, G., G. Dosi and G. Fagiolo (2002), 'On the ubiquitous nature of agglomeration economies and their diverse determinants: some notes', in A. Curzio and M. Fortis (eds), *Complexity and Industrial Clusters*, Heidelberg: Physica-Verlag.

Braczyk, H., P. Cooke and M. Heidenreich (eds) (1998), *Regional Innovation Systems*, London: UCL Press.

Brusco, S., G. Cainelli, F. Forni, M. Franchi, A. Malusardi and R. Righetti (1996), 'The evolution of industrial districts in Emilia-Romagna', in F. Cossentino, F. Pyke and W. Sengenberger (eds), *Local and Regional Response to Global Pressure: The Case of Italy and its Industrial Districts*, Geneva: International Institute for Labour Studies.

Cantner, U. and H. Graf (2006), 'The network of innovators in Jena: an application of social network analysis', *Research Policy*, **35**, 463–80.

Carlsson, B. (2007), 'Innovation systems: a survey of the literature from a Schumpeterian perspective', in H. Hanusch and A. Pyka (eds), *The Companion to Neo-Schumpeterian Economics*, Cheltenham, UK and Northampton, MA, USA: Edward Elgar.

Casson, M. (1995), *Entrepreneurship and Business Culture*, Aldershot, UK and Brookfield, USA: Edward Elgar.

Chadwick, G. (1971), *A Systems View of Planning*, Oxford: Pergamon.

Cooke, P. (1992), 'Regional innovation systems: competitive regulation in the new Europe', *Geoforum*, **23**, 365–82.

Cooke, P. (2002), *Knowledge Economies*, London: Routledge.

Cooke, P. (2008a), 'Knowledgeable regions, Jacobian clusters and green innovation', in S. Sacchetti and R. Sugden (eds), *Knowledge in the Development of Economies: Institutional Choices under Globalisation*, Cheltenham, UK and Northampton, MA, USA: Edward Elgar.

Cooke, P. (2008b), 'Good innovation systems practice: Vinnväxt and beyond', presentation to VINNOVA Vinnväxt Teams, Arlanda, Sweden, 22 January 2008.

Cooke, P. and R. Huggins (2003), 'High technology clustering in Cambridge', in F. Sforzi (ed.), *The Institutions of Local Development*, Aldershot: Ashgate.

Cooke, P. and K. Morgan (1998), *The Associational Economy*, Oxford: Oxford University Press.

Cooke, P., C. de Laurentis, F. Tödtling and M. Trippl (2007), *Regional Knowledge Economies*, Cheltenham, UK and Northampton, MA, USA: Edward Elgar.

Dei Ottati, G. (2009), 'An industrial district facing the challenges of globalisation: Prato today', *European Planning Studies*, **17**, 1817–35.

Flyvbjerg, B. (2008), 'Curbing optimism bias and strategic *misrepresentation* in planning; reference class forecasting in practice', *European Planning Studies*, **16**, 3–22.

Goodall, C. (2007), *How to Live a Low-Carbon Life*, London: Earthscan.

Granovetter, M. (1985), 'Economic action and social structure: the problem of embeddedness', *American Journal of Sociology*, **91**, 481–510.

Hahn, O. (1984), *Equilibrium and Macroeconomics*, Oxford: Blackwell. (Original work published 1952)

Heiner, J. (1983), 'The origin of predictable behaviour', *American Economic Review*, **73**, 56–70.

Hirst, P. (1994), *Associative Democracy*, Amherst: University of Massachusetts Press.

Hodgson, G. (1993), *Economics and Evolution: Bringing Life Back Into Economics*, Cambridge: Polity Press.

Jacobs, J. (1969), *The Economy of Cities*, New York: Vintage.

Jones, D., J. Womack and D. Roos (1990), *The Machine that Changed the World*, London: Macmillan.

Jürgens, U., A. Blocker and S. MacNeill (2007), 'Automotive sector knowledge stocks and flows', Report to EU FP6 EURODITE Project, Berlin, WZB.

Knight, F. (1921), *Risk, Uncertainty and Profits*, Chicago: Chicago University Press.

Kristensen, P. (1992), 'Industrial districts in West Jutland, Denmark', in F. Pyke and W. Sengenberger (eds), *Industrial Districts and Local Economic Development*, Geneva: International Institute for Labour Studies.

Krugman, P. (1995), *Development, Geography and Economic Theory*, Cambridge, MA: MIT Press.

Landes, D. (1998), *The Wealth and Poverty of Nations*, London: Abacus.

Lundvall, B. and B. Johnson (1994), 'The learning economy', *Journal of Industry Studies*, **1**, 23–42.

McLoughlin, B. (1969), *Urban and Regional Planning: A Systems Approach*, London: Faber.

Maillat, D. (1995), 'Territorial dynamic, innovative milieux and regional policy', *Entrepreneurship and Regional Development*, **7**, 157–65.

Marshall, A. (1919), *Industry and Trade*, London: Macmillan.

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

Paniccia, I. (2006), 'Cutting through the chaos: towards a new typology of industrial districts and clusters', in B. Asheim, P. Cooke and R. Martin (eds), *Clusters and Regional Development*, London: Routledge.

Pavitt, K. (1984), 'Sectoral patterns of technical change: towards a taxonomy and a theory', *Research Policy*, **13**, 343–75.

Penrose, E. (1995), *The Theory of the Growth of the Firm*, Oxford: Oxford University Press. (Original work published 1959)

Romer, P. (1990), 'Endogenous technical change', *Journal of Economic Literature*, **98**, 338–54.

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

Sabel, C. (1995), 'Experimental regionalism and the dilemmas of regional economic policy in Europe', in *Proceedings of Local Systems of Small Firms Seminar*, Paris: OECD.

Schumpeter, J. (1975), *Capitalism, Socialism and Democracy*, New York: Harper Torchbooks.

Simon, H. (1962), 'The architecture of complexity', *Proceedings of the American Philosophical Society*, **106**, 181–201.

Simon, H. (1972), 'Theories of bounded rationality', in C. McGuire and R. Radner (eds), *Decisions and Organisations*, Amsterdam: North-Holland.

Varaldo, R. and L. Ferrucci (1996), 'The evolutionary nature of the firm within industrial districts', *European Planning Studies*, **4**, 16–23.

Vernon, R. (1966), 'International investment and international trade in the product cycle', *Quarterly Journal of Economics*, **80**, 190–207.

PART 3

NETWORK EVOLUTION AND GEOGRAPHY

12 Clusters, networks and economic development: an evolutionary economics perspective

Elisa Giuliani

1. Introduction

This chapter explores the relationship between networks and economic development in geographical clusters of firms,[1] by developing a conceptual framework based on evolutionary economics (Dosi, 1988; Nelson and Winter, 1982). This is an important issue in economic geography because industrial clusters have come to play an increasingly central role in the analysis of economic growth and competitiveness of countries (Krugman, 1991; Porter, 1998) and in the debate between convergent versus divergent economic growth (Martin and Sunley, 1998). In a context of increased trade liberalisation, in fact, processes of Marshallian (1920) industrial localisation are believed to fuel the emergence of 'growth poles' or successful clusters (Porter, 1998). This is often associated with the fact that firms operating in clusters are likely to generate a socio-economic environment, characterised by dense inter-firm networks, which enhances their likelihood to innovate.

The importance of networks in regional clusters has been emphasised by several scholars since the beginning of the 1980s (e.g. Camagni, 1991; Piore and Sabel, 1984; Scott, 1988). Part of these studies has focused on the analysis of social and business relations among economic actors located within clusters and has given emphasis to the role of the territory, which, via its localised networks, influences the innovative behaviour and performance of firms.[2] This focus became predominant in the 1990s, up to the point that, in a recent contribution, Boggs and Rantisi (2003) argued that economic geography, as a field of studies, has undertaken a 'relational turn', which pays more attention to linkages than to geography per se. The relational turn implies a micro-centred approach to analysis, which looks at firms and their interrelations and emphasises that 'economic actors themselves produce their regional environments' (Bathelt and Glückler, 2003, p. 123). Thus, during the 1990s, the emergence of successful clusters or districts became increasingly associated with the presence of localised networks, and with a local institutional thickness that lowers transaction costs and favours the diffusion of knowledge (e.g. Lawson and Lorenz, 1999; Rabellotti, 1999; Saxenian, 1994).

The relational turn in economic geography has introduced two types of question, which are, generally speaking, still open for investigation and can be fruitfully explored via an evolutionary approach. The first one has to do with the understanding of the (causal) relationship between the behaviour of individual actors – such as firms – and the networks they are capable of generating at the local level.[3] It is now a commonly acknowledged fact that firms are heterogeneous in their internal capabilities (Dosi, 1988) and their (relational) behaviour may shape the evolution of a territory (on this, see e.g. Lazerson and Lorenzoni, 1999; Markusen, 2003; Martin and Sunley, 2003; Maskell, 2001); as suggested by Bathelt and Glückler (2003) in fact 'sometimes, even one large dominant firm might cause regional growth or decline simply through its linkages with other regional agents' (pp. 121–2). Empirically, this has been shown by several studies (see e.g. Albino et al., 1999; Eraydin and Fingleton, 2006). However, the way firm characteristics and inter-firm heterogeneity affect the structural characteristics of networks is still a rather under-theorised topic in economic geography.

A second question relates to the nature of networks and their impact on the economic development of regional clusters. While economic geographers have given considerable emphasis to the importance of networks for economic development (e.g. Becattini, 1990; Camagni, 1991; Yeung, 2000) – as noted before, with few exceptions (e.g. see the works of Glückler, Chapter 13 in this book; Breschi and Lissoni, 2003, and Breschi et al., Chapter 16 in this book), very limited analysis has been carried out so far on the characteristics of these networks – in terms of their content and structure – and very little is known about how different structural properties of localised networks may affect economic development processes (on this aspect, see also the contribution of Cantner and Graf – Chapter 17 of this book). Implicit in at least part of the cluster literature is that *dense* networks, based on, for example, intense inter-firm cooperation, are desirable to achieve certain development or competitiveness goals (Dimitriadis et al., 2005; Nadvi, 1999; OECD, 1996; Pyke et al., 1990; Schmitz, 1995).[4] In light of this, policies have been designed to strengthen the formation of inter-firm linkages at the regional level (OECD, 2001). However, density is only a rough indicator of structure. Organisational sociologists (Burt, 1992; Gulati, 1998; Smith-Doerr and Powell, 2003) and, more recently, economists (Cowan and Jonard, 2004; Lavezzi, 2003) have looked at other dimensions of a network structure (see e.g. works on small worlds and scale-free networks) and have provided promising theoretical arguments and empirical evidence that lead one to believe that the structural properties of networks may influence the quality of economic development in regions.

In Giuliani and Bell (2005) and Giuliani (2007a), I have made an attempt to look at one dimension of the first question, exploring the relationship between the knowledge base of firms and the structural properties of knowledge networks in wine clusters. In this chapter I explore the second question mentioned above. In particular, I analyse the effect of two types of network – the business and the knowledge network – on the performance of firms in three wine clusters. Evolutionary economics constitutes an appropriate intellectual environment for exploring these questions, for the following reasons. It has an interest in micro-founding its theories, that is, trying to explain aggregate patterns of behaviour by the behaviour of its micro-level agents. As suggested by Dosi (1988), in fact, 'theories . . . must involve or at least be consistent with a story of what agents do and why they do it' (Dosi, 1997, p. 1531). Because firms are agents with bounded rationality, they experience imperfect learning paths that generate persistent heterogeneity among them, both in a widely considered economy, and, understandably, also within a regional cluster. Consequently, evolutionary economics views firms as agents that compete in a selective environment, and it considers selection to be the result of different historical paths of accumulation of knowledge in firms:

> while (imperfect) adaptation and discovery generate variety . . . collective interactions within and outside market, perform as *selection mechanisms*, yielding also differential growth (and possibly disappearance) of different entities which are so to speak 'carriers' of diverse technologies, routines, strategies. (Dosi, 1997, p. 1531)

This chapter is well embedded in the evolutionary economics literature because it considers firm heterogeneity and selectivity as elements that strongly condition the nature of development in regional clusters. In this sense it is intrinsically different from previous works in this field, where scholars have tended to emphasise the 'collective' nature of inter-firm interactions in clusters.[5] This study finds that knowledge networks, which are formed by more selective mechanisms than business networks, tend to have strong but unevenly distributed effects on the performance of cluster firms.

The chapter is organised as follows: section 2 elaborates the conceptual framework. Section 3 presents the data and the methodology of analysis. Section 4 reports and discusses the empirical results and finally section 5 concludes and provides suggestions for further research.

2. What type of networks for what type of development?

Networks: an overview of the literature

> Networks contribute significantly to learning processes, provide contexts for innovation, create knowledge spillovers, facilitate information transfers between small and large-scale firms, and provide social infrastructure to accelerate technological change and industrial innovation. Studies have demonstrated the role and importance of networks in a variety of economic contexts. (Murphy, 2003, p. 176)

Starting from the 1930s, the concept of networks has occupied a prominent place in diverse fields of research such as anthropology, psychology, sociology and molecular biology. In the field of organisational behaviour, Roethlisberger and Dickson (1939) described and emphasised the importance of informal networks in organisations, and network analysis has become an academic discipline after the influential work of Harrison White in the 1970s (e.g. White, 1970, 1992). Social networks are defined as 'a set of nodes (e.g. persons, organisations, etc.) linked by a set of social relationships (e.g. friendship, transfer of funds, overlapping membership of a specific type)' (Laumann et al., 1978, p. 458).

In the past two decades, networks have been increasingly studied in economics and related fields (e.g. regional studies, innovation studies, etc.), as also highlighted by Smith-Doerr and Powell (2003). The 'relational turn' in economic geography does itself centre around the concept of networks. At the end of the 1980s, following from Piore and Sabel's (1984) flexible specialisation theory, networks became incorporated into the various definitions of industrial agglomerations. For example, in 1988, Scott defined industrial localities as:

> agglomerations [of producers] that coalesce out of the dense networks of transactional interrelations that form as the social division of labour deepens and as particular groups of producers are brought into intense and many-sided interaction with one another. (p. 31)

In a similar vein, Camagni (1991) defined the innovative milieu as 'the set, or the complex network of mainly informal social relationships in a limited geographical area, [. . ..], which enhance the local innovative capability through synergetic and collective learning processes' (p. 3).

The embeddedness of firms in local networks has often been considered to be a way through which transaction costs are reduced, because networks breed trustworthy relations among firms (Granovetter, 1985). Because of geographic proximity, firms in clusters or industrial districts are often described as being embedded in local business networks, defined as 'an integrated and co-ordinated set of ongoing economic and non-economic relations embedded within, among and outside business firms' (Keeble and Wilkinson, 1999, p. 299). Business networks are important also because they enhance the diffusion of knowledge and the generation of local spillovers (Audretsch and Feldman, 1996; Caniëls and Verspagen, 2001; Simmie, 2003). A reason why knowledge spillovers are believed to be highly localised is that a relevant part of the knowledge that is transferred between firms has a tacit component, which is informally transmitted by face-to-face interactions.

As mentioned, economic geographers and cluster scholars have tended to give relevance to the *density* of networks (Amin and Thrift, 1992; Camagni, 1991; Garofoli, 1991; Keeble et al., 1999; Nadvi, 1999; Schmitz, 1995; Yeung, 2000), but still limited analysis has been carried out on how different types of network – in term of their content and especially of their structure – affect the emergence of successful and vibrant clusters. This is largely because of the fact that disentangling different networks' effects on performance is a difficult exercise, which may also require a shift in the method of analysis (Boggs and Rantisi, 2003; Markusen, 2003).

Recent works have advanced in this direction, applying *social network analysis* (Wasserman and Faust, 1994) to look at whether the structural positions of firms within different types of intra-cluster network affect their performance (see Bell, 2005; Boschma and Ter Wal, 2007; Giuliani, 2007b). In these works, the interest of the authors was to understand whether more central firms in different types of intra-cluster network achieved higher performances. In this chapter, instead, the interest is to explore whether the existence of a linkage between any dyad of cluster firms increases the probability of them both being 'good performers', that is, of achieving similarly high performances. This is elaborated in the following section.

Business networks, knowledge networks and economic development

In this chapter I explore the impact of two different intra-cluster networks (i.e. the business and the knowledge networks) on the likelihood that cluster firms perform similarly well. Thus, the interest here is not to explore whether a firm performs better than another by way of its better structural positioning, but to look at whether the presence of a given linkage between any two firms in the cluster affects the likelihood that these two firms are both good performers. This is not done with the objective of identifying the factors that lead the firms to be good performers but just to compare the relative power of two different networks in affecting this process.[6] As explained later in this section, this is done with the aim of investigating (1) which of the two networks influences most the performance of firms and, above all, (2) whether they fuel processes of even (or uneven) economic development within the cluster.

In order to carry out this analysis, I draw on a previous work (Giuliani, 2007a) carried out on three wine clusters (Colline Pisane and Bolgheri/Val di Cornia in Italy and Valle de Colchagua in Chile), in which I have compared the structural characteristics of two types of network: the business and the knowledge networks. The former is defined as the set of relationships established by firms in a cluster when they interact on issues related to their business. According to Giuliani (2007a), 'the formation of business networks is based on the coexistence of market, social and institutional relationships, which occur almost routinely in a cluster context' (p. 145). Examples of such interactions that give rise to the formation of a business network are the trade of inputs or services, membership in the same local consortium, or meeting at local industry events, which imply a personal direct interaction regarding, for example, firms' productive activities, the local labour market, international markets, and so on. A business interaction occurs also when two firms borrow each other's machinery or tools for production, or their technical employees meet and discuss their appropriate use or, finally, when firms buy each other's grapes or bulk wine, or when entrepreneurs gather together to fund a new *oenotourisme*, or wine tourism, initiative in the area.

The knowledge network is defined as the network that links firms through the transfer of innovation-related knowledge, aimed at the solution of complex technical problems.[7] The knowledge network thus is based on the transfer of knowledge among firms, which occurs informally for problem-solving and is promoted by the local community of technicians and entrepreneurs.

Giuliani (2007a) shows that business networks and knowledge networks are structurally different. The difference lies in the fact that, whereas business networks are pervasive, connecting, in a fairly homogeneous way, almost the entire population of cluster firms, knowledge networks are very selective, not only because they are less dense, but also because the linkages are unevenly distributed across the network. Figures 12.1 to 12.3 show these differences in the three clusters.

Giuliani (2007a) provides an interpretation of these observed differences. The shape of the business networks is considered to depend on the serendipity through which these types of linkage are formed in clusters. These are places where firms operate in the same industry and at close geographical proximity, and therefore it is conceivable that entrepreneurs and employees are embedded into a tightly-knitted social space (as in e.g. Becattini, 1990), which favours trustful connections. Thus, business interactions are favoured by the existence of firms' geographical, sectoral and social proximity in the cluster (Boschma, 2005). This suggests that the business network is likely to be shaped by pervasive and unplanned local interactions – quite in line with Pyke at al. (1990), Saxenian (1994) and Malmberg (2003). As mentioned by Malmberg (2003), in fact, 'local interactions are characterised not just by being unstructured and unplanned, but also relatively broad and diffuse, sometimes unwanted and often seemingly of little immediate use' (p. 157). Similarly, Saxenian (1994) describes the informal conversations among engineers in Silicon Valley as 'pervasive' (p. 33).

While it is understandable, within the most accepted framework, to envisage business interactions as pervasive, less straightforward is the understanding of the selective nature of knowledge networks.[8] Evolutionary economics represents a natural environment in which to interpret this latter result. It does so, first, as it provides the theoretical background that explains why firms are persistently heterogeneous (Nelson and Winter, 1982). Second, it relates the role of path dependency and firm heterogeneity to the processes of adoption of innovations, and diffusion of knowledge (Arthur, 1988; Dosi, 1991). Firms are heterogeneous in their internal capabilities as their learning processes are highly idiosyncratic and path-dependent, such that 'past technological achievements

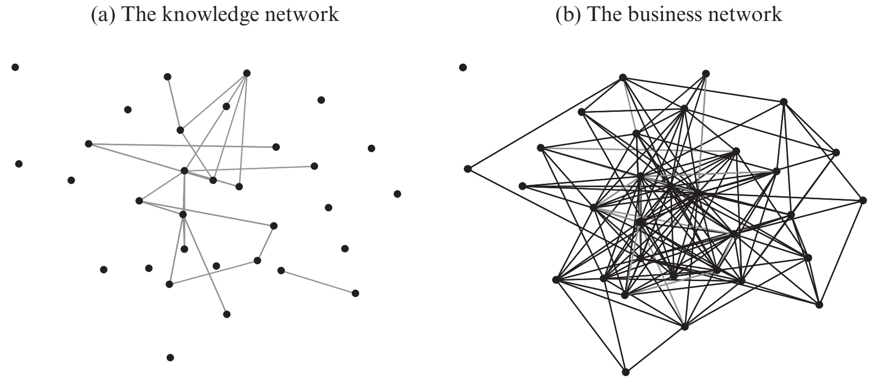


Figure 12.1 The knowledge network and business network in Colline Pisane

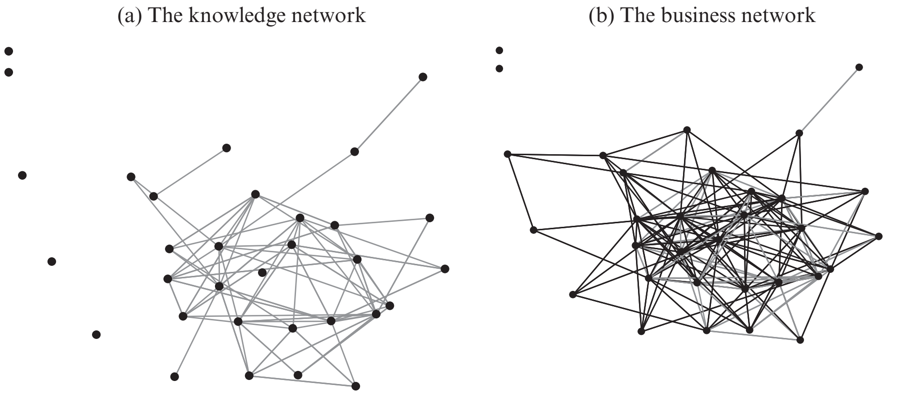


Figure 12.2 The knowledge network and business network in Valle de Colchagua

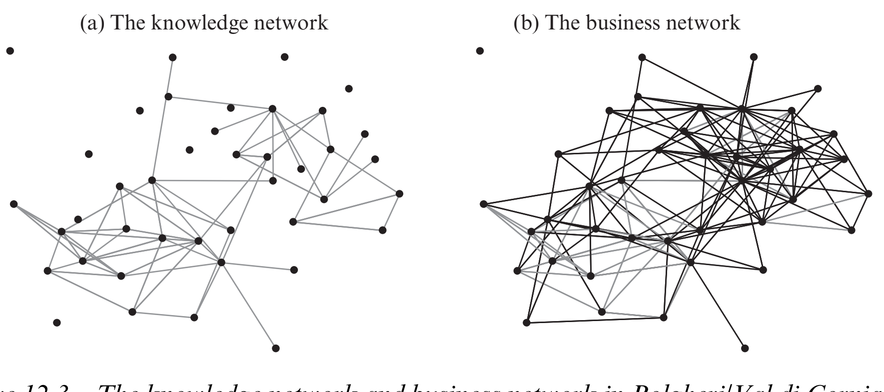


Figure 12.3 The knowledge network and business network in Bolgheri/Val di Cornia

influence future achievements via the specificity of knowledge that they entail' (Dosi, 1991, p. 183). This heterogeneity is in turn considered to influence the way innovations are diffused in an economy. As Dosi (1991) put it:

> A whole approach to innovation and diffusion studies would agree that it is often the case that the adopting firm differ in technological capabilities, and that some of the potential adopters may not adopt because they do not have the technological and organizational capabilities to do so. To put it simply, they do not adopt since they lack of appropriate skills, internal knowledge, or managerial capabilities. (p. 187)

It is thus suggested that what influences the pattern of diffusion of an innovation is the 'nature and distribution of technological asymmetries between firms' (Dosi, 1991, p. 187), and therefore their different capacities to absorb or make profitable use of a given technology at a given point in time (Cohen and Levinthal, 1990). In line with this, Giuliani (2007a) finds that the process of formation of knowledge networks in the three wine clusters analysed is based on the relative strength of the cluster firms' knowledge bases.

In particular, the study finds that firms with particularly strong knowledge bases are likely to be perceived by other cluster firms as 'technological leaders' in the local area, leading to them being sought out as sources of advice and knowledge more often than firms with weaker knowledge bases (see also Giuliani and Bell, 2005). Furthermore, firms with stronger knowledge bases have higher absorptive capacity and therefore have more incentives to search for external knowledge, as they know that they will be able to make profitable use of it. By the same token, firms with strong knowledge bases are more likely targeted by those cluster firms whose 'cognitive distance' from the technological leaders is not too high to inhibit communication (Boschma, 2005). A consequence of this is that firms with similarly strong knowledge bases do exchange knowledge more intensively than firms with weak knowledge bases. From an economic viewpoint, firms with stronger knowledge bases have incentives to transfer knowledge to other organisations when these have equally advanced knowledge bases and are therefore in a position to reciprocate with valuable knowledge. In line with von Hippel (1987) and Schrader (1991), in fact, reciprocation constitutes the expected pay-off for the transferred knowledge.

Consequently, the structural characteristics of business and knowledge networks vary because they are grounded on differing underlying rationales. Given these differences, an interesting question arises about which of the two networks affects the likelihood that cluster firms perform similarly well. This question has some relevance for the following reasons (see Table 12.1 for an overview). At a very first approximation, one could argue that if business networks were a powerful channel for the diffusion of 'benefits' at the local level (i.e. spillovers), these would be pervasively spread within the cluster, giving cluster firms a similar chance to benefit from those spillovers and to improve their performance accordingly. Such a mechanism could progressively smooth down the differences between cluster firms' performance, promoting processes of more even economic development at the local level, while increasing the differences with other areas (consistent with the argument that clustered firms grow more or perform better than isolated firms). In contrast, one could argue that, if knowledge networks were a powerful channel to enhance, for example, innovation, the benefits of the transfer of knowledge

Table 12.1 Expected relationship between network structure and performance

Type of networks	Network structure	Suggested determinants of structure	Expected impact on economic development
Business network	Pervasive	Meso-level forces (e.g. geographical/sectoral/ social proximity)	Progressively more even at the local level
Knowledge network	Selective	Micro-level forces (e.g. firm knowledge bases or other internal characteristics)	Progressively more uneven at the local level

would be spread selectively in the cluster, driving persistent heterogeneity among firm performance. In this case, the success of a cluster would be related to a much smaller number of firms and not to the entire community of firms. In light of this, the emergence of successful clusters could depend on the behaviour of individual firms – quite in line with evolutionary thinking.

3. Methodology

Context and data

This empirical study is contextualised in the wine industry. In the recent two decades, wine consumption has dramatically changed, shifting market preferences from quantity, non-premium wines to quality, premium wines. On the side of production, technology and techniques of grape-growing and wine making have undergone processes of increased codification of knowledge. Technical change has been strong in the industry and the key competitive asset of wine producers is now the capacity to absorb and manage new techniques of production.

This study has been carried out in two countries, Italy and Chile, which differ historically but have recently undergone a similar process of wine industry growth and modernisation. In both cases, what has sparked growth is a process of technological change aimed at improving the quality of wines. Based on this, new and successful wine clusters have developed in both countries since the 1980s. This study considers two clusters in Italy (Colline Pisane and Bolgheri/Val di Cornia) and one in Chile (Valle de Colchagua). The boundaries of these wine clusters are given by their natural conditions. These types of cluster are therefore easily identifiable economic entities, whose boundaries are nowadays set by the Denomination of Origin regulations applied internationally by wine-producing countries. All the three clusters are territories densely populated by fine wine producers and by grape growers. The degree of vertical division of labour is however rather low, with no other relevant suppliers localised within the clusters' territory. On a global scale, these clusters can be classified as 'followers', with Bolgheri/Val di Cornia and Valle de Colchagua being more dynamic than Colline Pisane.

This study is based on micro-level data, collected at the firm level in three clusters in years 2002–03. The analysis has required careful data collection through face-to-face interviews. Interviews were carried out with the skilled workers (i.e. oenologists or agronomists) and the survey was directed to producers of fine wines. Data were gathered using the universe of fine wine producers populating the three clusters, comprising 32 in Colline Pisane, 41 in Bolgheri/ValdiCornia and 32 in Valle de Colchagua, summing up to a total of 105 firms. The data were collected using a structured questionnaire, which allowed relational data to be collected through a 'roster recall' method so that each firm was presented with a complete list (roster) of the other firms in the cluster, and was asked to report about their relations with other firms (see the Method section).

Method

This chapter adopts multiple regression quadratic assignment procedure (MRQAP), originally developed by Mantel (1967) and Krackhardt (1987), to test whether business and knowledge networks influence a similarity matrix of firm performance, which is used here as a proxy of the evenness in the performance of cluster firms. This method considers the general model:

$$Y = \alpha + \beta_1 X_1 + \beta_2 X_2 + \varepsilon$$

where Y is a square (network) matrix of dependent observations, X_1 is a square matrix of observations on the independent variable of interest, and X_2 is a square matrix of observations on a second independent variable of interest, α is a constant term, and ε is a square matrix of residuals. The MRQAP evaluates the significance of β_1 and β_2 in a different way from conventional econometrics. Since the problem with statistical analysis of social network data is the lack of independence of dyadic observations (Laumann and Pappi, 1976), the MRQAP permutes the data many times, re-estimating the β_1 and β_2 coefficients under each permutation, and then comparing the observed β_1 and β_2 coefficients to the null hypothesis reference distribution created by the set of re-estimated β_1 and β_2. The specific method adopted in this chapter is based on double Dekker semi-partialling procedure, developed by Dekker et al. (2005) and performed through UCINET 6.116.

Dependent variable The dependent variable is represented by a similarity squared matrix of firm performance. The matrix is composed of n rows and n columns, corresponding to the number n of firms in the network. Each firm has a row and a column, which are labelled 1, 2, . . . n. Each cell in the matrix reports the existence of a similarity existing in the performance of firm i in the row and firm j in the column. In this specific case, in the matrix, there is a 1 in the (i, j) cell if firm i and firm j are *both* good performer firms. Cells (i, j) report a 0 when either both firms are poor performers or at least one of the two is. The matrix that results from this is called *Good performer*. A firm is considered here a 'good performer' if either of two criteria are met: (1) one of the firm's wines has been rated as having good quality in the international wine journal *Wine Spectator* in the period 2003–05; (2) the firm's export share is higher than the cluster average.

The first indicator (1) adopted to measure the good performance of firms (*Rating*) is drawn from *Wine Spectator*.[9] The wine rating of this journal is based on the quality assessment of an international panel of expert oenologists, who review more than 12,000 wines each year in blind tasting. After tasting, oenologists assign a score to each wine brand according to a 100-point scale, ranging from 100, when the wine is of outstanding quality, to 50 when it is of poor quality. A set of information is listed with each rated wine: the wine vintage, the wine area and the market price. This indicator is valued 1 when any of the firm's wines has been assigned at least 70 points in years 2003–05, the minimum threshold for a wine to be considered of drinkable quality and to be recommended by the journal. It is valued 0 otherwise. A lag of two years is allowed between the year in which the interviews were carried out and the vintage of the most recent rated wines.[10]

The second indicator (2) is a measure of the export propensity of firms (*Export*) and it tries to correct the fact that about 70 per cent of firms were not rated under (1), especially in the Colline Pisane cluster where only 12 per cent of the firms were rated. Since the objective of this chapter is not to explore the performance of firms in absolute terms (i.e. compared with competing firms on international markets), but to assess the degree to which firms diverge in their performance relative to the cluster, it seemed appropriate to consider other dimensions of performance. The share of exports, compared to total production, is a very rough approximation of performance. However, qualitative evidence collected in the field suggests that highly exporting firms are normally considered in this sector to be more successful.

Independent variables The first independent variable is the *knowledge network*. This network refers only to knowledge that is transferred or received with the specific purpose of solving technical problems. The questions formulated to collect knowledge network data are reported below:

Q1 If you are in a critical situation and need technical advice, to which of the local firms mentioned in the roster do you turn?
[Please rate the importance you attach to the knowledge linkage established with each of the firms according to its persistence and quality, on the basis of the following scale: 0 = none; 1 = low; 2 = medium; 3 = high].

Q2 Which of the following firms do you think have benefited from technical support from this firm?
[Please rate the importance you attach to the knowledge linkage established with each of the firms according to its persistence and quality, on the basis of the following scale: 0 = none; 1 = low; 2 = medium; 3 = high].

These questions specifically address problem solving and technical assistance because they involve some effort in producing improvements and change within the economic activity of a firm. This is meant to go beyond the mere transfer of information, whose access can be easily attained through other channels (e.g. trade fairs, the internet, specialised reviews etc.). Instead, the interest here is to investigate whether local stocks of contextualised complex knowledge are not only accessible but also eventually absorbed by localised firms. So, for example, knowledge is transferred by providing a suggestion on how to treat a new pest or how to deal with high levels of wine acidity during fermentation. Accordingly, the knowledge transferred is normally the reply to a query on a complex problem that has emerged and that the firm seeks to solve.

These questions have been used to construct an $n \times n$ matrix. Each cell in the matrix reports the existence of knowledge being transferred from firm i in the row to firm j in the column. I use dichotomous data so in the matrix there is a 1 in the (i, j) cell if firm i has transferred knowledge to firm j. Cell (i, j) contains a 0 when no transfer of knowledge has been reported to occur between firm i and j. This is a directed matrix.

The second independent variable is the *business network*. The business network includes all market and non-market relations established by wine producers in the development of their productive activities. Similarly to the question above, relational data about firms' interactions was collected through a roster recall method, so that each firm was presented with a complete list (roster) of the other firms in the cluster, and was asked the question reported below:[11]

> *Q3* With which of the cluster firms mentioned in the roster do you interact for business matters?
> [Please indicate the frequency of interaction according to the following scale: 0 = none; 1 = low; 2 = medium; 3 = high]

Question Q3 has been used to construct an $n \times n$ matrix. Each cell in the matrix reports the existence of business interactions between firm *i* in the row and firm *j* in the column. I use dichotomous data so in the matrix there is a 1 in the (*i*, *j*) cell if firm *i* and firm *j* have interacted for business matters. Cell (*i*, *j*) contains a 0 when no business interactions have been reported to occur between firms *i* and *j*. This is an undirected matrix.

For the sake of simplicity, control variables are not introduced in the MRQAP model. This represents a limitation of the current study. However, it should be noted that this chapter is not looking for factors that affect similarity in firms' performance, but is comparing the relative importance of two different types of network on this aspect. Moreover, a previous work (Giuliani, 2007b), based on standard econometrics, and including a number of firm-level control variables, achieves results that are consistent with those found in this chapter (section 4).

The analysis of data also includes some descriptive statistics about the structural properties of the business and knowledge networks. The measures adopted for this purpose are explained in the Appendix.

4. Empirical results

As shown by Table 12.2, in Colline Pisane the number of good performers is more limited than in the other two clusters, as only 12.5 per cent of the firms are rated by *Wine Spectator* and 34 per cent qualify as good exporters. Given the fact that in this specific case all the firms that are rated are also good exporters, only 34 per cent of the firms can be considered good performers. The cluster of Valle de Colchagua is at odds with Colline Pisane, since about 70 per cent of the firms are good performers, with 60 per cent being rated by *Wine Spectator*. Finally, Bolgheri/Val di Cornia is situated somewhat in the middle, with about 50 per cent of the firms being classified as good performers.

Table 12.3 reports the results of the QAP multiple regression, showing that in all three cases the β for the knowledge network has a positive and significant sign, while no significant evidence is found for the business network. This means that the existence of a knowledge linkage between any two firms increases the likelihood that both are good performers. In contrast, the existence of a business linkage is not associated with a higher similarity in the good performance of firms.

In the remainder of this section I will try to provide an interpretation of these results. In particular, I will link the data on firm performance presented in Table 12.2 with some of the structural properties of the knowledge networks (Table 12.4), which had an influence on the similarity matrix of firm performance (*Good performer*) (Table 12.3). The first

Table 12.2 Cluster firms' performance indicators

Indicators of firm performance	Colline Pisane (%)	Bolgheri/Val di Cornia (%)	Valle de Colchagua (%)
Rating	12.5	22	60
Export	34	34	53
Good performer (Rating + Export)	34	49	72

Table 12.3 Results of the MRQAP: the impact of knowledge and business networks on the good performance of firms

Independent *(Good performer)*	Colline Pisane	Bolgheri/Val di Cornia	Valle de Colchagua
Intercept	0.0946	0.1206	0.4522
Knowledge network	0.277***	0.156**	0.215***
Business network	0.021	−0.002	0.121
R-square	0.029	0.010	0.041
N. permutations	2000	2000	2000

Notes: *** significant at 1%; ** significant at 5%.

cluster, Colline Pisane, has two properties. The first one is the relatively high presence of low performers (Table 12.2); the second is the poor density and high fragmentation of its knowledge network.[12] Table 12.4 reports some descriptive statistics of this cluster's knowledge network, showing that it has the lowest network density, if compared with the other two clusters (0.04). More interestingly, it has the highest levels in the Hirschman/Herfindahl (HH) score applied to a measure both of actor centrality (*coreness*) (0.311) and of network fragmentation (0.919 and 0.756). This indicates that the distribution of knowledge linkages in the cluster is more uneven than in the other two clusters and the vast majority of firms are disconnected from the network.

The cluster of Valle de Colchagua shows a quite opposite situation. As shown by Table 12.2 it has a higher incidence of good performers (70 per cent). It is striking that, if compared to the other two clusters, it has a relatively higher inclusion firms in the intra-cluster knowledge network. This is measured by the several indicators listed in Table 12.4. In more detail, the HH indicator is the lowest among the three (0.046), indicating a relatively less uneven distribution of knowledge linkages between cluster firms. Similarly, the measures of network fragmentation are the lowest among the three clusters (0.690 and 0.442). This seems consistent with the fact that the highest incidence of good performers is matched with the higher participation of firms in the intra-cluster knowledge network.

Interestingly enough, the case of Bolgheri/Val di Cornia is in line with the results found for Colline Pisane and Valle de Colchagua as it is situated between the two, both in terms of the incidence of good performers (Table 12.2) and of the inclusion of firms in the intra-cluster knowledge network (Table 12.4).

These results seem therefore to suggest that the structural properties of these knowl-

Table 12.4 Structural indicators of knowledge and business networks

	Colline Pisane	Bolgheri /Val di Cornia	Valle de Colchagua
Knowledge network			
Density	0.04	0.05	0.09
Hirschman/Herfindahl on actor coreness	0.311	0.091	0.046
Fragmentation – reciprocated ties only	0.919	0.878	0.690
Fragmentation – unreciprocated ties also	0.756	0.395	0.442
Business network			
Density	0.32	0.20	0.30
Hirschman/Herfindahl on actor coreness	0.010	0.014	0.012
Fragmentation	0.063	0.049	0.123

edge networks may affect the degree of unevenness of cluster firm performance. It is striking that the presence of local business networks is instead not found to have an impact on this. As Table 12.4 shows, the business networks are structurally similar across the three clusters and, overall, they are denser and less fragmented than the knowledge networks.

On the basis of these results, it is possible to consider that, at least in these three clusters, the presence of pervasive and relatively dense business networks does not translate into benefits for the community of firms located in the cluster. It seems plausible to argue that this may be a result of the *content* of the ties of this network, determined by the qualitative nature of the relationships (Gulati, 1998; Rodan and Galunic, 2004). In spite of their pervasiveness, business linkage may not be channels for the transfer of valuable knowledge or be the vehicle of any other pecuniary advantage. For example, belonging to the same business association or interacting in a trade fair may spark new ideas and opportunities when firms already have internal capabilities to take advantage of the interaction. Otherwise, the potential advantage of an interaction is left in a vacuum. This clearly indicates that the characteristics and behaviour of firms play a central role in determining the beneficial effects of network embeddedness.

For this reason, knowledge networks appear to have a more powerful impact on the performance of firms. The selectivity through which they are formed suggests that firms with stronger knowledge bases and more advanced absorptive capacity are more likely to participate in the local knowledge network (Giuliani, 2007a). This in turn may guarantee a certain degree of quality in the *content* of the knowledge transferred, because this type of firm is more likely to transfer high quality knowledge about, for example, frontier methods of production. The downside of this is that the selectivity of the knowledge network may signify higher returns/performance exclusively for the firms that are connected to the knowledge network and no effect on those cluster firms that are either excluded or peripheral in the network. By way of this, firms that are centrally positioned in the cluster's knowledge networks may improve their position and performance in a self-reinforcing way, becoming a sort of local 'elite', and widening the gap with more isolated or marginalised firms. Through the selective development of knowledge networks, thus, the heterogeneity in the distribution of firm knowledge bases may breed persistent unevenness among cluster firms, leading their performance to diverge over time. In this way, moreover, the process of economic development of a cluster may become vulnerable and dependent on the behaviour of individual firms, rather than on any other meso-level force.

5. Conclusions

Geography has come to play an increasingly central role in the analysis of economic growth and competitiveness of countries (Krugman, 1991; Porter, 1998). This is considered to be because of the fact that, in the context of increased trade liberalisation, processes of Marshallian industrial localisation fuel the emergence of 'growth poles' or successful clusters. This is often associated with the fact that firms operating in clusters are likely to generate a social environment, characterised by dense inter-firm networks, which enhances their likelihood to innovate. In this chapter I have argued that two types of question are still open for investigation in this domain of study. The first one has to do with the understanding of the causal relationship between the behaviour of individual actors – such as firms – and the structural properties of the networks that they are capable of generating at the local level. The second one relates to the relationship between the characteristics of localised networks – in terms of their content and structure – and the economic development of regions.

This chapter deals with the second question. In particular, it has analysed the effect of two types of network – the business and the knowledge network – on the similarity in the performance of firms in three wine clusters. The analysis has been carried out using methods of *social network analysis* (Wasserman and Faust, 1994). Multiple regression quadratic assignment procedure (Dekker et al., 2005) is applied to test the impact of the knowledge and business networks on the dependent variable represented by a similarity squared matrix of firm performance.

The results of the QAP multiple regression show that in all three clusters the knowledge network has a positive and significant impact on the dependent variable, while no significant evidence is found for the business network. This implies that the existence of a knowledge linkage between any two firms increases the likelihood that both firms are good performers – which is not the case for the business linkages. These results raise two types of consideration. First, that the *content* of the network ties may be extremely important for the economic performance of firms and that it is not networking per se that enhances performance in clusters, but the existence of valuable, knowledge-rich linkages. Second, that the *structure* of knowledge-rich networks may indeed affect the *quality* of regional economic development: where the knowledge network was slightly less selective – as in Valle de Colchagua – the number of good performers was higher. The key issue is to understand what determines the formation of valuable or knowledge-rich linkages and what is the relative weight of firm-level variables and territory-specific factors.

My contention is that this is very much affected by the characteristics of firms in clusters (see also Lazerson and Lorenzoni, 1999; Martin and Sunley, 2003; Maskell, 2001) – an issue that is in line with an evolutionary economics view of the firm. As I am able to show in Giuliani (2007a), in fact, the structure of knowledge-rich networks is shaped by the distribution of firms' knowledge bases. According to this, it is possible to suggest that localised (knowledge-rich) networks can even amplify or facilitate the sharing of valuable resources, but the seeds of success are primarily in the internal activities of cluster firms. This is nevertheless an issue that deserves further research, as I am suggesting in the rest of this section.

This analysis was set within specific empirical and methodological limits. The first is that this is a single industry study. The generalisation of its results is therefore bounded by the specificities of the wine industry. The second is that, for the sake of simplicity, the MRQAP analysis does not introduce control variables in the model, as discussed in section 3. The third limitation is that this study focuses only on intra-cluster linkages, leaving aside extra-cluster linkages, which are instead an important component of firms' learning processes (Amin and Thrift, 1992; Bell and Albu, 1999; Coe and Bunnell, 2003). Finally, a further limitation is particularly important and raises interesting questions: the cross-sectional nature of this analysis. In this respect, future research should explore the evolution of networks in time, and look at the existence of micro–meso interactions. In particular, it would be interesting to answer questions like: what types of initial micro-level conditions are necessary to spark the formation of (knowledge-rich) networks that lead to equitable forms of economic development in regional clusters? What network structures facilitate the emergence of vibrant clusters? Thus, future research should try to explore the interaction between several types of firm-level characteristic and their likelihood to engage in networks – including social/friendship networks – that can enhance their economic performance and sustain processes of economic development. For example, one could ask: under what firm-level conditions do social linkages favour the formation of knowledge linkages and generate a less selective environment in regional clusters, than the one observed in this study? In sum, it would be extremely valuable to explore the relationship between the distribution of given firm-level characteristics (internal capabilities, degree of foreignness, size, strategic orientation, etc.) in a territory, the structural evolution of different types of network (knowledge, business, friendship, etc.) and the patterns of economic development.

To conclude, it is hoped that, in spite of the existing limitations, this chapter serves to spark a discussion on the need to study networks in geography (e.g. in clusters or districts) with more analytical rigour than has been achieved in earlier studies. While most of the cluster literature and, to some extent, policy makers, (over)emphasise the importance of networks, still very little is known about which type of network is relevant for economic development and which type, instead, generates just a local 'buzz' with no tangible outcome. Also, very little is known about what micro-level mechanisms lead to the evolution of networks that can promote processes of economic development. Evolutionary economics seems to me to constitute the right conceptual approach through which these issues can fruitfully be explored in the future.

Notes

1. A 'cluster' is defined here as a geographical agglomeration of firms operating in the same industry, in line with the definition given by Humphrey and Schmitz (1996), among others.
2. In this respect, Camagni (1991) argues that 'a link-up with a firm located in Silicon Valley is more a link with the Valley itself than with a special firm, with which if otherwise located, no agreement would probably be made' (p. 140).
3. In effect, most of the literature has paid attention more to the effects of networks on firms, than the reverse – as suggested in Giuliani and Bell (2005).

4. At the opposite end, scholars have noted that too dense networks and tight forms of social control may cause undesirable effects in the economy of a cluster (Gordon and McCann, 2005).
5. See, for example, works on 'collective learning' (see *Regional Studies*, **33** (4), 1999) and 'collective efficiency' (see *World Development*, **27** (9), 1999).
6 For this reason, this chapter performs an analysis of multiplexity, which tests the existence of a statistical relationship between sets of networks composed of the same nodes (Lazega and Pattison, 1999). This allows the effects of networks to be studied considering their structural properties, rather than the characteristics of the individual firms (see section 3).
7. The definition of innovation-related knowledge transfer used in this chapter is fairly simplistic. However, pilot interviews for this study have revealed that the knowledge transfer aimed at the solution of technical problems is a predominant and very important part of the inter-firm knowledge diffusion process in wine clusters. This in turn may be considered a relevant source of innovation for firms, as indicated by von Hippel (1987).
8. Note that selectivity entails not only a lower density of linkages, but also an uneven distribution of them. Thus, whereas it is conceivable that the restricted way of defining knowledge linkages in this work may influence their lower density, the same does not necessarily apply to their distribution.
9. The choice of the journal was made on the basis of two criteria: international respectability in the wine industry, free availability on the web and ample coverage (countries, vintages, wine areas).
10. It is worth noting that the majority of the wines tasted are submitted to *Wine Spectator* by the wineries or their US importers. Additionally, the journal spends substantial effort in buying and reviewing wines that are not submitted, at all price levels. Accordingly, a firm's wines may not be rated for three main reasons: first, because of a self-selection effect, firms with poor quality achievements will have little incentive to send their wines to the journal for assessment. Second, US importers will not recommend and signal wines to *Wine Spectator* when they consider them of poor quality. Third, *Wine Spectator* itself selects out all the wineries producing very poor quality wines. These considerations suggest therefore that firms whose wines are not rated tend to be poor performers. The same applies for firms whose wines are rated but are assigned less than 70 points.
11. When asked this question, professionals were requested to mention only firms with which any linkage was formed for a business-related matter, independent of the underlying reasons that led to the formation of that linkage – for example whether it was based on the existence of a solid friendship between the parties or on purely arm's-length relations. For the sake of simplicity in data collection, this question does not allow market-based relationships to be disentangled from socio-institutional relationships.
12. Density refers to the number of linkages present in a network, calculated on the theoretical maximum, while fragmentation is a measure that counts the number of pairs of nodes that are disconnected from each other. See also the Appendix for details on these measures.

References

Albino, V., C. Garavelli and G. Schiuma (1999), 'Knowledge transfer and inter-firm relationships in industrial districts: the role of the leader firm', *Technovation*, **19**, 53–63.

Amin, A. and N. Thrift (1992), 'Neo-Marshallian nodes in global networks', *International Journal of Urban and Regional Research*, **16** (4), 571–87.

Arthur, W.B. (1988), 'Competing technologies: an overview', in G. Dosi, C. Freeman, R. Nelson, G. Silverberg and L. Soete (eds), *Technical Change and Economic Theory*, London: Pinter, pp. 590–608.

Audretsch, D. and M.P. Feldman (1996), 'R&D spillovers and the geography of innovation and production', *American Economic Review*, **86** (3), 630–40.

Bathelt, H. and J. Glückler (2003), 'Toward a relational economic geography', *Journal of Economic Geography*, **3**, 117–44.

Becattini, G. (1990), 'The Marshallian industrial district as a socio-economic notion', in F. Pyke, G. Becattini, W. Sengenberger (eds), *Industrial Districts and Inter-Firm Co-operation in Italy*, Geneva: International Institute for Labour Studies, pp. 37–51.

Bell, G. (2005), 'Clusters, networks, and firm innovativeness', *Strategic Management Journal*, **26**, 287–95.

Bell, M. and M. Albu (1999), 'Knowledge systems and technological dynamism in industrial clusters in developing countries', *World Development*, **27** (9), 1715–34.

Boggs, J.S. and N.M. Rantisi (2003), 'The "relational turn" in economic geography', *Journal of Economic Geography*, **3**, 109–16.

Borgatti, S.P. and M.G. Everett (1999), 'Models of core/periphery structures', *Social Networks*, **21**, 375–95.

Borgatti, S.P. (2006), 'Identifying sets of key players in a network', *Computational, Mathematical and Organizational Theory*, **12** (1), 21–34.

Boschma, R.A. (2005), 'Proximity and innovation: a critical assessment', *Regional Studies*, **39** (1), 61–74.

Boschma, R.A. and A.L.J. Ter Wal (2007), 'Knowledge networks and innovatiove performance in an industrial district. The case of the footwear district in the South of Italy', *Industry and Innovation*, **14** (2), 177–99.

Breschi, S. and F. Lissoni (2003), 'Mobility and social networks: localised knowledge spillovers revisited', CESPRI Working Papers 142, CESPRI, Centre for Research on Innovation and Internationalisation, Università Bocconi, Milano, Italy.

Burt, R.S. (1992), *Structural Holes: The Social Structure of Competition*, Cambridge, MA: Harvard University Press.

Camagni, R. (1991), *Innovation Networks: Spatial Perspectives*, 1st edn, London, New York: Belhaven Press.

Caniëls, M.C.J. and B. Verspagen (2001), 'Barriers to knowledge spillovers and regional convergence in an evolutionary model', *Journal of Evolutionary Economics*, **11**, 307–29.

Coe, N.M. and T.G. Bunnell (2003), 'Spatialising knowledge communities: towards a conceptualisation of transnational innovation networks', *Global Networks*, **3** (4), 437–56.

Cohen, W.M. and D.A. Levinthal (1990), 'Absorptive-capacity – a new perspective on learning and innovation', *Administrative Science Quarterly*, **35** (1), 128–52.

Cowan, R., and N. Jonard (2004), 'Network structure and the diffusion of knowledge', *Journal of Economic Dynamics and Control*, **28**, 1557–75.

Dekker, D., D. Krackhardt and T.A.B. Snijders (2005), 'Sensitivity of MRQAP tests to collinearity and autocorrelation conditions', Working Paper Series on Research in Relationship Management, Nijmegen School of Management, University of Nijmegen.

Dimitriadis, N., M. Simpson, A. Andronikidis (2005), 'Knowledge diffusion in localised economies of SMEs: the role of local supporting organisations', *Environment and Planning* C, **23** (6), 799–814.

Dosi, G. (1988), 'Sources, procedures and microeconomic effects of innovation', *Journal of Economic Literature*, **26** (3), 1120–71.

Dosi, G. (1991), 'The research on innovation diffusion: an assessment', in N. Nakicenovic and A. Grubler (eds), *Diffusion of Technologies and Social Behavior*, Berlin: Springer-Verlag, pp. 179–208.

Dosi, G. (1997), 'Opportunities, incentives and the collective patterns of technological change', *The Economic Journal*, **107**, 1530–47.

Eraydin, A. and D. Fingleton (2006), 'Network relations and local economic development: some causes of differentiated network structures and intensities among Turkish industrial firms', *Environment and Planning A*, **38**, 1171–86.

Garofoli, G. (1991), 'Local networks, innovation and policy in Italian industrial districts', in E.M. Bergman, G. Maier and F. Tvdtling (eds), *Regions Reconsidered. Economic Networks, Innovation and Local Development in Industrialised Countries*, London: Cassel, pp. 119–140.

Giuliani, E. (2007a), 'The selective nature of knowledge networks in clusters: evidence from the wine industry', *Journal of Economic Geography*, **7**, 139–68.

Giuliani, E. (2007b), 'Networks and heterogeneous performance of cluster firms', in K. Frenken (ed.), *Applied Evolutionary Economics and Economic Geography*, Cheltenham, UK and Northampton, MA, USA: Edward Elgar, pp. 161–89.

Giuliani, E. and M. Bell (2005), 'The micro-determinants of meso-level learning and innovation: evidence from a Chilean wine cluster', *Research Policy*, **34** (1), 47–68.

Gordon, I.R. and P. McCann (2005), 'Innovation, agglomeration, and regional development', *Journal of Economic Geography*, **5**, 523–43.

Granovetter, M. (1985), 'Economic action and social structure: the problem of embeddedness', *American Journal of Sociology*, **91** (3), 481–510.

Gulati, R. (1998), 'Alliances and networks', *Strategic Management Journal*, **19**, 293–317.

Humphrey, J. and H. Schmitz (1996), 'The triple C approach to local industrial policy', *World Development*, **24** (12), 1859–77.

Keeble, D., C. Lawson, B. Moore and F. Wilkinson (1999), 'Collective learning processes, networking and "institutional thickness" in the Cambridge region', *Regional Studies*, **33** (4), 319–32.

Keeble, D. and F. Wilkinson (1999), 'Collective learning and knowledge development in the evolution of regional clusters of high technology SMEs in Europe', *Regional Studies*, **33** (4), 295–303.

Krackhardt, D. (1987), 'QAP partialling as a test of spuriousness', *Social Networks*, **9**, 171–86.

Krugman, P. (1991), *Geography and Trade*, Cambridge, MA: MIT Press.

Laumann, E.O., J. Galaskiewicz and P.V. Marsden (1978), 'Community structure as interorganizational linkages', *Annual Review of Sociology*, **4**, 455–84.

Laumann, E.O. and F.U. Pappi (1976), *Networks of Collective Action: A Perspective on Community Influence Systems*, New York: Academic Press.

Lavezzi, A.M. (2003), 'Complex dynamics in a simple model of economic specialization', Discussion Paper, Dipartimento di Scienze Economiche University of Pisa, 2.

Lawson, C. and E. Lorenz (1999), 'Collective learning, tacit knowledge and regional innovative capacity', *Regional Studies*, **33** (4), 305–17.
Lazega, E. and E.P. Pattison (1999), 'Multiplexity, generalized exchange and cooperation in organizations: a case study', *Social Networks*, **21**, 67–90.
Lazerson, M.H. and G. Lorenzoni (1999), 'The firms that feed industrial districts: a return to the Italian source', *Industrial and Corporate Change*, **8**, 235–66.
Malmberg, A. (2003), 'Beyond the cluster – local milieus and global economic connections', in J. Peck, H.W.C. Yeung (eds), *Remaking the Global Economy*, London: Sage, pp. 145–59.
Mantel, N. (1967), 'The detection of disease clustering and a generalized regression approach', *Cancer Research*, **27**, 209–20.
Markusen, A. (2003), 'Fuzzy concepts, scanty evidence, policy distance: the case for rigour and policy relevance in critical regional studies', *Regional Studies*, **37**, 701–17.
Marshall, A. (1920), *Principles of Economics*, London: MacMillan.
Martin, R. and P. Sunley (1998), 'Slow convergence? The new endogenous growth theory and regional development', *Economic Geography*, **74** (3), 201–27.
Martin, R. and P. Sunley (2003), 'Deconstructing clusters: chaotic concept or policy panacea?', *Journal of Economic Geography*, **3**, 5–35.
Maskell, P. (2001), 'The firm in economic geography', *Economic Geography*, **77** (4), 329–44.
Murphy, J.T. (2003), 'Social space and industrial development in East Africa: deconstructing the logics of industry networks in Mwanza, Tanzania', *Journal of Economic Geography*, **3**, 173–89.
Nadvi, K. (1999), 'Shifting ties: social networks in the surgical instrument cluster of Sialkot, Pakistan', *Development and Change*, **30** (1), 141–75.
Nelson, R.R. and S.G. Winter (1982), *An Evolutionary Theory of Economic Change*, Cambridge, MA: Belknap Press of Harvard University Press.
OECD (1996), *Networks of Enterprises and Local Development: Competing and Co-operating in Local Productive Systems*, Paris: OECD.
OECD (2001), *Innovative Clusters: Drivers of National Innovation Systems*, Paris: OECD.
Piore, M.J. and C.F. Sabel (1984), *The Second Industrial Divide: Possibility for Prosperity*, New York: Basic Books.
Porter, M. (1998), 'Clusters and the new economics of competition', *Harvard Business Review*, November–December, 77–90.
Pyke, F., G. Becattini and W. Sengenberger (1990), *Industrial Districts and Inter-Firm Co-operation in Italy*, Geneva: International Institute for Labour Studies.
Rabellotti, R. (1999), 'Recovery of a Mexican cluster: devaluation bonanza or collective efficiency?', *World Development*, **27**, 1571–85.
Rodan, S. and D.C. Galunic (2004), 'More than network structure: how knowledge heterogeneity influences managerial performance and innovativeness', *Strategic Management Journal*, **25**, 541–62.
Roethlisberger, F.J. and W.J. Dickson (1939), *Management and the Worker*, Cambridge, MA: Harvard University Press.
Saxenian, A. (1994), *Regional Advantage Culture and Competition in Silicon Valley and Route 128*, Cambridge, MA: Harvard University Press.
Schmitz, H. (1995), 'Small shoemakers and Fordist giants: tale of a supercluster', *World Development*, **23** (1), 9–28.
Schrader, S. (1991), 'Informal technology transfer between firms: cooperation through informal trading', *Research Policy*, **20**, 153–70.
Scott, A.J. (1988), *New Industrial Spaces. Flexible Production Organization and Regional Development in North America and Western Europe*, London: Pion Limited.
Simmie, J. (2003), 'Innovation and urban regions as national and international nodes for the transfer and sharing of knowledge', *Regional Studies*, **37** (6 & 7), 607–20.
Smith-Doerr, L. and W.W. Powell (2003), 'Networks and economic life', in N. Smelser and R. Swedberg (eds), *The Handbook of Economic Sociology*, New York: Russell Sage Foundation and Princeton University Press.
von Hippel, E. (1987), 'Cooperation between rivals: informal know-how trading', *Research Policy*, **16**, 291–302.
Wasserman, S. and K. Faust (1994), *Social Network Analysis. Methods and Applications*, Cambridge: Cambridge University Press.
White, H.C. (1970), *Chains of Opportunity*, Cambridge, MA: Harvard University Press.
White , H.C. (1992), *Identity and Control: A Structural Theory of Social Action*, Princeton, NJ: Princeton University Press.
Yeung, H.W.-C. (2000), 'Organizing "the firm" in industrial geography I: networks, institutions and regional development', *Progress in Human Geography*, **24** (2), 301–15.

Appendix: Measures of network structure

1. Density

Network density is defined as 'a measure of the proportion of possible lines that are actually present in a graph' (Wassermann and Faust, 1994, p. 101). It is calculated as follows: the maximum possible number of lines in a graph is given by the sum of possible unordered pairs of nodes: $n(n - 1)/2$ with g being the number of nodes in the network. The density (DN) is calculated as the ratio of the number of lines present, L, to the maximum possible:

$$DN = (L/(n(n - 1)/2))$$

2. Actor coreness

Actor 'coreness' is defined as the degree of closeness of each node to a core of densely connected nodes observable in the network, as described by Borgatti and Everett (1999).

3. Fragmentation

Fragmentation is a measure that counts the number of pairs of nodes that are disconnected from each other. The measure is found in Borgatti (2006) and is calculated as follows: given a matrix R in which $\mathrm{r}ij = 1$ if i can reach j and $\mathrm{r}ij = 0$ otherwise, network fragmentation is:

$$F = 1 - (2\sum\sum \mathrm{r}ij/n(n - 1))$$

The indicator ranges from a minimum of 0 to a maximum of 1.

13 Reputation, trust and relational centrality in local networks: an evolutionary geography perspective*

Stefano Denicolai, Antonella Zucchella and Gabriele Cioccarelli

1. Introduction: social variety in an evolutionary economic geography

The literature about clusters and districts highlights that geographical proximity facilitates knowledge sharing and, therefore, innovation processes (Breschi and Lissoni, 2001; Cooke, 1998). According to a more or less tacit assumption in this stream, local knowledge is somewhat 'in the air': all district firms apparently have equal access to this important asset. This perspective tends to underestimate the relevance and the evolutionary path of knowledge creation *inside* the firm, which generates a variety of routines: each organization is approached by researchers as a sort of 'black box', while the focus goes on the relationships among local agents (Boschma and Kloosterman, 2005). Moreover, economic geography pays attention to whether generic locations or local networks are relevant for the development and sustainability of a competitive advantage (Boschma, 2004; Castells, 1996). This leads to the assumption – according to the traditional stream – that, in a particular cluster, local firms share similar levels of absorptive capacity. This circumstance provides a unique and inimitable location-bounded completive advantage for any firm located in a given place.

Evolutionary economic geography adds a complementary view to these assumptions. It highlights that, also within the same district, firms could be very heterogeneous in their knowledge and competence bases (Boschma and Ter Wal, 2007; Nelson and Winter, 1982). The variety of firms and their evolution over time are important dimensions of analysis. Organizations involved in local networks aren't 'black boxes': they differently absorb, explore and exploit local knowledge, because they use different routines, beliefs and habits.

Knowledge flows are a critical issue in this framework. Geographical proximity facilitates knowledge sharing and, thus, interactive learning and innovation. However, local firms show cognitive schemes that depend on both know-how and social dynamics (Bathelt and Glückler, 2003; Cohen and Levinthal, 1990; Nooteboom, 2004). These two dimensions, especially within local clusters, are deeply intertwined (Granovetter, 1985). Firms' absorptive capacity is grounded, first of all, on knowledge bases, but it significantly relies also on social issues such as culture, trust and reputation (Cohen and Levinthal, 1990; Malmberg and Maskell, 2002). Knowledge externalities in districts are not necessarily 'in the air', because knowledge tends to remain inside the boundaries of firms and of local networks (Zucchella et al., 2004). Since local agents differ in terms of absorptive capacity, the above mentioned accumulation of knowledge and social assets increases firms' variety over time. *The critical issue addressed in this chapter is that the* *distribution and the evolution of firm variety within districts are also influenced by the local social capital*. The latter is a local asset that is non-tradable and historically developed following a path-dependent evolution, representing an interesting field of research in the evolutionary economics perspective (Nelson and Winter, 1982).[1] According to the evolutionary economic geography perspective, both firm- and network-specific characteristics are considered (micro and meso levels of analysis).

Economic geography tries to explain the concentration and distribution of economic activities across space and has already taken into account the critical role of social capital (Dei Ottati, 1994; Putnam, 1993). To integrate an evolutionary approach means to consider the processes that have determined these social patterns (Boschma and Frenken, 2006). This perspective is particularly coherent with the nature of factors such as trust and reputation. As mentioned before, these social assets aren't tradable and can't be modified radically in the short term (Zucchella and Ganzaroli, 2003). They are the outcome of a long path-dependent process, stemming from a series of mutual interactions, which changes the cognitive schemes and beliefs of local inhabitants and firms (Nooteboom, 2004). In this sense, economic geography, economic growth and social dynamics are strictly interconnected (Boschma and Frenken, 2006). In order to manage social assets, firms develop a specific know-how stored into organizational routines, rituals and habits. This is a valuable tacit knowledge that is not homogeneously spread in the territory: social assets represent an important dimension of firm variety. According to these assumptions, a better explanation for the mutual relationship between firm- and network-specific social resources becomes an important research issue in evolutionary economic geography.

The core research issues of this chapter are represented by trust and reputation inside local clusters. Trust refers to local- and network-specific resources, because it encompasses at least two agents. Second, reputation refers to firm-specific intangible resources and permits us to analyse the social variety within the district. Indeed, it is reasonable to assume that, within local clusters, a firm's reputation is grounded also on its superior ability in the management of social relationships (Ely and Valimaki, 2003; Nooteboom, 2002).

In this conceptual framework, the development of trust in each dyadic relation (network resources) shapes the reputation degree (firms' intangible resources) of the actors involved, thus altering the degree of variety of social capital within the district. Second, the evolution of reputation influences trust development at the network level. Over time, this loop creates a pattern of mutual interactions in the business activity where 'pieces' of knowledge are embedded in both social ties and single organizations (Brown and Duguid, 1991; Morgan, 1997). This close connection between trust and reputation has also been confirmed by other studies (Aringhieri et al., 2006; Dahl and Pedersen, 2005; Elangovan and Shapiro, 1998; Xiong and Liu, 2003).

After this introduction, in the second section, this contribution proposes an operationalization of 'trust' and 'reputation' and a short overview about the related literature. The third section focuses on the main links between reputation and network theory, with special attention to its impact on inter-organizational structure, leading us to formulate two research propositions. The latter are tested in the fourth section, discussing the results obtained from a network analysis on three Italian tourism local clusters, where a large number of small sized firms (hotels, ski schools, travel agencies and so on) compete and cooperate in providing a wide range of services to the final consumer. Finally, in the fifth section, by using the described theoretical and empirical issues, some conclusions are drawn.

2. Trust and reputation: an integrated and evolutionary perspective

The impact of social capital on district dynamics and local network structure has been considered in a large number of studies (Cioccarelli, 2003; Elangovan and Shapiro, 1998; Granovetter, 1985; Gulati, 1995; Nahapiet and Goshal, 1998; Nooteboom et al., 1995). In order to evaluate the impact of social capital – with a special emphasis on the issues of trust and reputation – on firm variety within districts, according to an evolutionary geography perspective, two research steps are necessary. First of all, the effective reciprocal influence between social assets, such as trust, and intangible firm-specific resources, such as reputation, must be confirmed. Moreover, in this attempt, it is also important to verify if network structure plays a role as a sort of 'mediator' between pure externalities available to all firms and resources absorbed by different local agents. Network relationships influence perceptions, information distribution and, above all, the effective access to knowledge externalities (Gulati, 1999). Then, as a second research step, it is important to understand 'how' – in a longitudinal perspective – social capital, at the network level, shapes social resources at the firm level, selecting the best know-how and increasing heterogeneity. This chapter focuses on the first of the above mentioned research steps – analysis of the correlation between network- and firm-specific social heterogeneity – as a fundamental precondition in order to approach later the evolutionary side of the problem.

Trust and reputation can mitigate information asymmetry by reducing transaction-specific risks (Xiong and Liu, 2003) and coordination costs (Granovetter, 1985; Perrone, 2001), strengthening the relationships, increasing the effectiveness of network structure and empowering the ability of local firms to explore and exploit location-specific assets. On the other hand, trust and reputation reinforce the consolidation of organizational routines (Nelson and Winter, 1982; Winter, 2002) increasing their uniqueness and inimitability, while geographic proximity facilitates the development of informal inter-organizational patterns of interactions (Denicolai and Zucchella, 2005).

In order to analyse social capital at the 'meso-level' – district/network relationships – we consider the construct of trust. This issue has been of growing interest in the economics literature in the last decade, opening a lively debate about this concept, its relevance for economic analysis and its theoretical modelling (Gambetta, 1998; Gulati, 1995; Nooteboom, 2002; Ring and Van de Ven, 1994). These studies have proposed different perspectives of analysis and have generated different operationalizations.. The definition of trust adopted in this contribution is the following:

> the belief that an individual or group makes good-faith efforts to behave in accordance with any commitments both explicit and implicit, is honest in whatever negotiations preceded such commitment, and does not take excessive advantage of another even when the opportunity is available. (Cummings and Bromiley, 1996, p. 303)

The analysis of this definition shows that the object of this informal agreement or the nature of this expectation can be differentiated. Many authors emphasize the complexity of trust issues and draw a distinction between different forms (Barber, 1983;

Gulati, 1995; McAllister 1995; Nooteboom, 2002; Shapiro, 1987; Zucker, 1986). Making distinctions between different types of trust is important, because any breach of trust affects network structure, generating different dynamics and, hence, requires different responses (Nooteboom, 2002). In consideration of the assumptions of our research and the specificity of the selected empirical field, among the various forms of trust, our study distinguishes between relational trust and competence-based trust.

Relational trust is related to aims, intention, dedication/care, leadership capability, benevolence and so on. It emphasizes the social dimension of the relationship. In particular, the degree of relational trust is closely connected to the past experiences shared between local actors and is path dependent (Rousseau et al., 1998). A positive past experience can strengthen the relation and reinforce the development of inter-firm routines, while a negative one can decrease the degree of relational trust between the parties. Relational trust supports, above all, the tacit knowledge transfer among local firms (Hansen, 1999). In this perspective, trust is a sort of evolutionary mechanism with a 'memory' that stores knowledge about common past experiences. The balance between the positive and the negative experiences generates trust or distrust and thus impacts on the selection process of partners and routines over time (Essletzbichler and Rigby, 2007). The concept of 'positive' or 'negative' experience may be vague: cognitive patterns and absorptive capacities of different firms can sometimes evaluate the same experience differently. Therefore, this form of trust includes many individual or emotional factors that influence the absorptive capacity of the firm in a recursive loop (Cohen and Levinthal, 1990).

The concept of *competence-based trust* relates to the perception of the effective contribution of the partners in terms of knowledge provided in the framework of the inter-organizational relationship. Collaborative know-how, embedded in the local inter-organizational routines, emerges only if all network agents provide adequate know-how in order to achieve the network aims, and if relative absorptive capacities permit its effective exploitation/exploration (Nooteboom, 2004; Simonin, 1997). Competence-based trust concerns elements such as ability, skills, languages, quality of products and services, financial credibility and so on. This type of trust is usually more 'rational' than the relational one, because it is often based on the consideration of more objective information, such as economic and financial indicators or product/service performance.

Second, in order to look at social capital at the 'micro-level' – firm variety – we consider 'reputation'. Many theories of the firm, including the resource-based view, the resource dependence stream, the transaction cost theory, agency and game theories, recognize the importance of reputation for organizations (Cao and Schniederjans, 2006; Dahl and Pedersen, 2005; Noorderhaven et al., 2002; Roberts and Dowling, 2000; Saxton, 1997). This topic concerns many disciplines: organization theory, psychology, sociology, marketing, finance, economics and so on. Reputation is often described in terms of customer loyalty, market trust, financial credibility, and name recognition. Moreover, a number of empirical surveys on reputation have already been published (e.g. Ely and Valimaki, 2003; Engelmann and Fischbacher, 2003). This heterogeneity has produced a vague and indistinct concept of reputation, and a substantial absence of a generally agreed definition in the literature. The Mori British Institute[2] defines corporate reputation as follows:

> the totality of emotional and intellectual disposition towards an organization that derives from the assumptions, perceptions and beliefs about what an organization is, how it is run and what it stands for.

This definition highlights how reputation is a non-tradable intangible resource stemming from a development process over time that depends on past events and choices. Moreover, reputation plays a significant role in the time–space distribution of routines through the so-called 'routine replication process' (Boschma and Frenken, 2006; Kogut and Zander, 1993). The latter is partial and imperfect, thus it leads to a contingent adaptation that stimulates variety (Boschma and Frenken, 2006).

An interesting study by Cambridge University (Brady, 2002) deepens the sources of this emotional and intellectual disposition and suggests that corporate reputation is grounded on seven factors:

1. emotional connections;
2. social credibility;
3. leadership, vision and desire;
4. knowledge and skills;
5. quality;
6. financial credibility;
7. environmental credibility.

The dynamics of these elements develop the firm reputation that is the result of a wide signalling activity (Fombrun and Shanley, 1990; Shapiro, 1983) based on available information on the above-mentioned factors. However, reputation is also shaped over time by the individual perception and interpretation of this signalling activity among firms, customers and other stakeholders (Fombrun, 2001; Kreps and Wilson, 1980). Three determining factors of reputation are very similar to the mentioned sources of 'relational trust': emotional connections (1), social credibility (2) and leadership (3). While three more elements seem closely connected to the concept of 'competence-based trust' and in general to more objective characteristics of the firm: knowledge and skills (4), quality (5), financial credibility (6). Environmental credibility (7) can be considered as a constant within local clusters, since it is connected to context-specific factors.

If different forms of trust exist, then reputation is also a complex phenomenon characterized by different forms (Cao and Schniederjans, 2006). Therefore, according to the taxonomy adopted in this study, we can introduce a distinction between *relational reputation* and *competence-based reputation*.

We expect that relational trust affects relational reputation (Aringhieri et al., 2006). It means that network structure influences firm variety. The degree of reputation is correlated to local social capital and past experiences known and shared by all local network nodes (Nooteboom and Gilsing, 2004) and especially concerns factors such as emotional connections, social credibility and leadership capability. A firm that informs another one about a positive [negative] business experience with a third firm, can increase [decrease] the reputation degree of the latter, especially when there is a strong trust relation between the two firms that have shared the information about that business experience (Köszegi, 2002; Zucker, 1986). This feeds the selection process of partners and routines. Trust and past experiences create a cognitive map of social identities and reputation about each

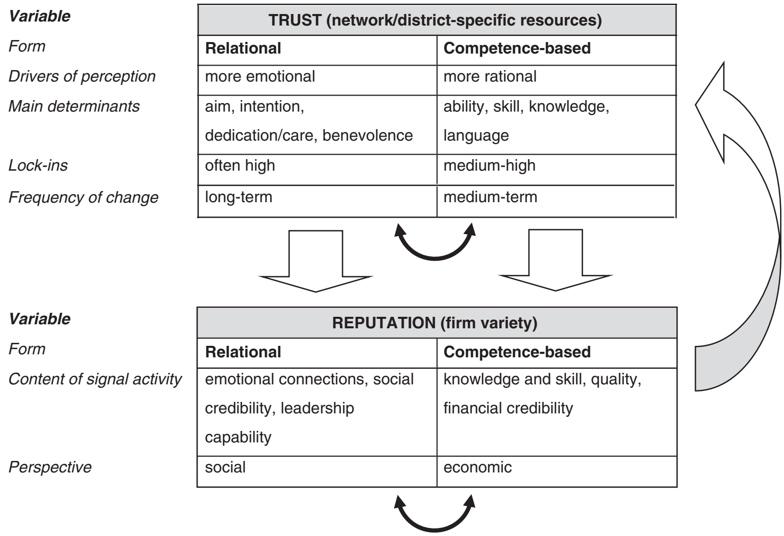


Figure 13.1 Dynamic relation between trust and reputation (evolutionary loop)

individual (Merry, 1984). When network members interact over a long period of time, they develop shared cognitive structures and inter-organizational routines (Denicolai and Zucchella, 2006; Nelson and Winter, 1982). Reputation may also reflect professional competence (Barber, 1983; Dobing, 1993; McKnight et al., 1995). In this case, competence-based reputation is influenced by competence-based trust, which involves factors such as knowledge and skills, quality of product/service, financial credibility.

Figure 13.1 shows a synthesis of the notions discussed in this section.

Trust dynamics in the local network is one of the most important sources of firm reputation and, at the same time, reputation can change the social network structure, especially in terms of the trust dimension. For example, a firm's good reputation can increase the expectation that its behaviour will be fair and predictable. In other words, reputation can increase the degree of trust between firms. A similar reciprocal interaction is observed between competence-based and relational trust, as well as between competence-based and relational reputation. Trust and reputation are two different social capital phenomena within districts, but there exists a self-influencing evolutionary loop between them. Over time, this process affects firm variety within the geographical cluster, especially in terms of social capital. Second, environmental events and district-specific path dependence influence the evolution of this loop by selecting the best inter-organizational dynamic processes among local agents, which then become relatively stable inter-organizational routines based also on social mechanisms.

3. Informal structure of local networks: the dynamics of nodes' centrality

The previous section highlighted the relationship between the district-specific social capital and the accumulation of social resources within the firms. It contributes to explain the heterogeneity of local agents. In order to complete this theoretical framework, a focus on the impact of network structure on firm reputation is necessary.

The literature emphasizes the importance of network structure (e.g. Gemunden et al., 1996; Van Haverbeke and Noorderhaven, 2001), the position of the firm inside the network (Gulati, 1999), the type of inter-organizational ties (Burt, 1992; Powell, 1990; Uzzi, 1996), the role of social dynamics (Donet and Cannon, 1997; Leana and Van Buren, 1999; Nooteboom, 2002), and knowledge sharing (Brock-Smith, 1997; Ritter, 1999) as factors of successful firm performance and evolution. Many scholars highlight how knowledge spillovers occur in social networks (Almeida and Kogut, 1999; Maggioni and Uberti, 2005). This chapter explores the circumstances in which network features, such as centrality or density degree, can alter the above described social evolutionary process, by supporting some firms located in given positions in their network. In particular, this section establishes a distinction between district- and network-specific social resources. Social networks do not necessary require permanent co-location for interactive learning to take place (Torre and Rallet, 2005). It's wrong to assume that all networks are geographically co-localized, even though geographical proximity and social interactions should significantly support collaborative initiatives (Boschma and Ter Wal, 2007; Bunnell and Coe, 2001).

Network value added evolves over time, becomes embedded in social dynamics, and concretely expresses its potential as inter-organizational routines (Foss, 1999; Gulati, 1999; Nelson, 1993; Saxenian, 1994; Simonin, 1997). In order to preserve the competitive advantages deriving from their ability to operate efficiently and flexibly as members of a community of connected firms, small firms located within a restricted geographical area must reduce the costs of their collaboration. This can be done within social relationships by using trust and reputation as an informal coordination system (Perrone, 2001).

The interest in local networks, such as districts and local clusters, has produced many theoretical studies (Garofoli, 1991; Marshall, 1920; Sforzi, 1990) and empirical surveys (Bellandi, 1989; Dei Ottati, 1994). In international literature, districts and local clusters seem to be more or less synonymous, but there are some differences, which an analysis of alternative definitions can confirm. By district we mean:

> a socio-territorial entity which is characterized by the active presence of both a community of people and a population of firms in one naturally and historically bounded area. (Becattini, 1990, p. 39)

According to this definition, four elements characterize the district; that is, firms' high specialization, a homogeneous system of values, strong influence of the social dimension, and geographical concentration. In this vision, territorial resources and patterns of interaction among firms and institutions clearly play a central role. On the other hand, according to a definition adopted in international literature, by local cluster we mean:

> geographically proximate firms in vertical and horizontal relationships involving a localized firm support infrastructure with shared developmental vision for business growth, based on competition and co-operation in a specific market field. (Cooke and Huggins, 2002, p. 52)

The core elements characterizing the notion of cluster are a stronger emphasis on economic rather than social factors and inter-firm dynamics, a shared vision for business growth, and no explicit mention of the role of the social dimension as the network 'glue'. Our study doesn't distinguish between these two concepts in the empirical survey. However, we report these definitions because it's important to underline that, in the context of this survey, social dynamics are always a fundamental dimension in the evolutionary economic geography perspective. In our opinion, the interconnectedness between knowledge flows and trust relationships can't be ignored in the study of evolution of routines within districts/clusters.

According to Glückler (2007), the relation between geography and networks can be studied in two main ways. First of all, proximity affects network formation. The association between district area and local network is not unidirectional: proximity does not constrain network formation. Networks should also embrace external agents, but social interactions should create a partial overlap between network and district boundaries (Storper and Walker, 1989). Second, the firm's position in the network structure makes a difference. This chapter focuses especially on this statement. In order to estimate the impact of network structure on the formation of patterns of social interactions and on reputation distribution, it is important to evaluate the position of each actor within the local cluster in terms of structural properties. Each actor can be studied by mapping the social network structure and adopting some indicators, such as relational centrality, closeness, inter-position degree, and so on (Kilduff and Tsai, 2003). On the other hand, network structural indicators are closely connected to other ones (Kilduff and Tsai, 2003; Krackhardt, 1990) and the parallel analysis of correlation with other factors – such as trust and reputation – is very complex and mutually interdependent. Our study focuses on the dimension of relational centrality, even though we are aware that other important dimensions exist, figured by alternative structural indicators. Having a good degree of centrality within the network means, for instance, having good access to critical information, which is an important source of organizational power (Daft, 2001).

Our study aims at adding more elements to this debate with a focus on the system of relations among three dimensions:

1. the centrality of nodes in the network structure;
2. the one-to-one trust relation;
3. the firm's reputation as defined above.

Within a local network, the stock of knowledge can be shared among many firms, creating a spread phenomenon capable of influencing the whole network, which can change the reputation of many firms involved in the network. Many authors argue that reputation is, above all, the result of complex interactions among firms (Dentchev and Heene, 2003; Fombrun, 2001). Based on the conceptual background outlined in section 2, all the dyadic trust relationships across local network nodes influence the reputation degree of the firms involved in the local network. The direct correlation between reputation and trust and their impact on inter-organizational structure is significant and confirmed by the literature (Elangovan and Shapiro, 1998; Ely and Valimaki, 2003). For example, the willingness to avoid gaining a bad reputation reduces opportunistic behaviour in the estimated long-period relationship between two actors.

Every node is analysed in terms of four variables: centrality, relational reputation, competence-based reputation, and economic performance, which allowed us to estimate a positive or negative impact of trust and reputation. All nodes can be linked with one another by informal relations: relational trust, competence-based trust and knowledge flows. The impact of a firm's both relational and competence-based reputation on network structure can be appreciated in terms of correlation with the centrality of the same firm inside the network structure.

This study aims at verifying this correlation, with reference to the following research propositions:

(Hp1) *There is a positive correlation between the degree of relational reputation and degree of centrality of a single actor inside the network.*

(Hp2) *There is a positive correlation between the degree of competence-based reputation and degree of centrality of a single actor inside the network.*

These kinds of correlation, if they exist, influence the network trajectory (Glückler, 2007; Kenis and Knoke 2002; Kilduff and Tsai 2003). The question is whether the structural dimensions of a social network and its variety at time 1 affect the interactions between local actors at time 2, and how this evolutionary path shapes patterns of knowledge sharing and firm heterogeneity. The network trajectory combines elements of evolutionary, network and economic geography streams. As discussed above, the next empirical section covers only the static side of the research question, and not the network trajectory dimension. However, the cumulative evolution associated with the network trajectory emphasizes the continuous processes of social embedding (Gulati and Gargiulo, 1999), which highlights the relevance of factors such as trust and reputation in an evolutionary economic geography framework.

4. An analysis of tourist districts: empirical findings

The theoretical sections indicate two patterns for the analysis of the impact of trust and reputation on network structure and inter-organizational routine formation. In particular, section 2 focuses on the meaning and features of the concepts of trust and reputation; section 3 approaches the problem by building on network theory. The aim of this section is to merge the two patterns, and to test the two research propositions formulated by using an empirical survey carried out on three Italian tourist districts interpreted as local networks: Bormio, Cortina d'Ampezzo and Pila. As the three local clusters share the same economic activity, findings can well be compared.

Bormio is an alpine town in the centre of high Valtellina, province of Sondrio, with approximately 4100 inhabitants, and a tourist tradition dating back to the second half of the nineteenth century. Its alpine valleys and passes have always been points of passage for people and products between northern and southern Europe. Cortina is a town located in the Dolomites with about 6500 inhabitants and a long tourist tradition. On the contrary, Pila, a small hamlet in the vicinity of Gressan (Aosta Valley), has a more recent tourist history and operates mostly during the winter season. Tourist districts are particularly interesting for this type of research because: (1) the firms involved operate in a complex system of symbiotic inter-dependencies; and (2) informal dynamics, such as knowledge sharing, trust and reputation play a very important role.

The empirical research was carried out through network analysis methodologies (George and Allen, 1989; Knoke and Kuklinski, 1982; Nieminen, 1974) and based on

Table 13.1 Survey statistics

Local cluster	Number of firms and local institutions (network node)	Questionnaires submitted	Questionnaires collected	Redemption rate (%)
Bormio	91	91	85	93
Cortina	61	61	30	49
Pila	50	50	24	48
Total	*202*	*202*	*139*	*69*

structural indicators and graphic network representation (George and Allen, 1989; Nieminen, 1974), calculated by different algorithms. A questionnaire was submitted to all the main local tourist actors, including firms – travel agencies, hotels, ski schools and so on – and local institutions (see Table 13.1).[3]

The questionnaire included a specific question for each variable of analysis: knowledge sharing, relational-based trust, competence-based trust.[4] The questionnaires were developed to map all the ties across network nodes. Each question followed the same protocol: a list of all nodes (local actors) was presented to the respondent as a list of all possible 'simple flag' answers (0 = no relation with the local actor; 1 = relationship with the local actor). These data were coded in symmetric adjacency matrices (Nieminen, 1974). The symmetry of these matrices was based on the notion of reciprocity: if node 'A' declared that it exchanged information or knowledge with node 'B', the relation was included in the dataset only if node 'B' also declared that it exchanged knowledge with node 'A'. This property is fundamental for efficient interaction when inter-organizational relations are based on informal agreements (Sutter and Kocher, 2004). The same criterion was used for data regarding the one-to-one trust relations (relational and competence-based), operationalized respectively according to the traits outlined in section 2. The net graphs in Figures 13.2, 13.3 and 13.4 show all network ties – in terms of knowledge sharing – among all nodes in the three considered clusters. These graphs reflect the results of questionnaires where the respondents were asked to identify all the local counterparts (other nodes in the network) with whom they regularly exchanged knowledge and business experience over the medium and long period.

The degree of centrality of each node has been calculated with the following algorithm (Nieminen, 1974):

$$c_j = \sqrt{\sum_{i=1}^{n} (I_{ji} * W_i)^2}$$

C_j = centrality of node '*j*' (where $0 < C_j < 1$)
I_{ji} = effective connection of node '*j*' with node '*i*' (1 = activated; 0 = not activated)
W_i = node's weight (where $0 < W_i < 1$)

Then, based on the theoretical assumptions mentioned in the previous sections, the degree of relational reputation (*RR*) and competence-based reputation (*CR*) of each node has been calculated as an aggregate of the trust relation of all network members in respect of that specific node:

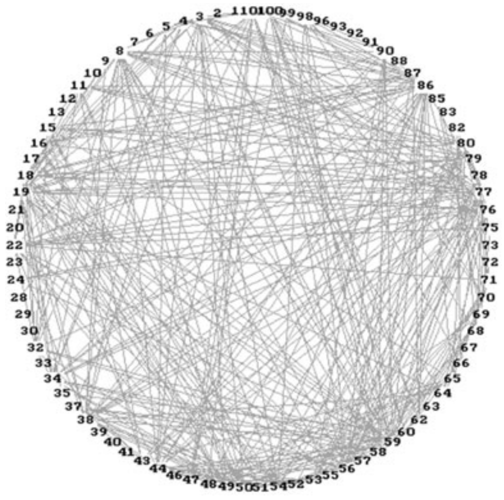


Figure 13.2 Knowledge sharing in Bormio

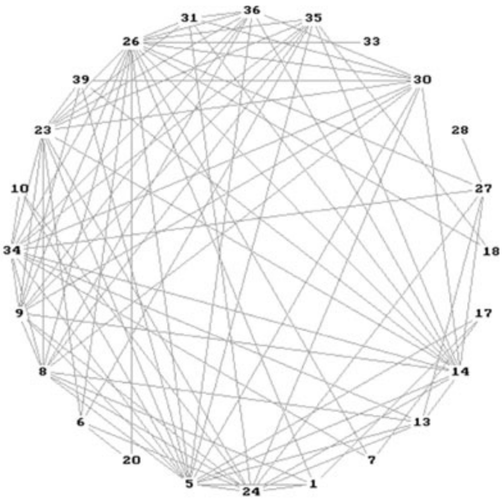


Figure 13.3 Knowledge sharing in Pila

$$RR_j = \frac{\sum_{i=1}^{n} (RT_{ji} * WR_i)}{\text{network size}}$$

RR_j = relational reputation of node 'j' (where $0 < RR_j < 1$)
RT_{ji} = effective relational trust relation of node 'j' with node 'i' (1 = activated; 0 = not activated)
WR_i = node's weight in terms of relational trust centrality (where $0 < WR_i < 1$)
network size = total number of nodes.

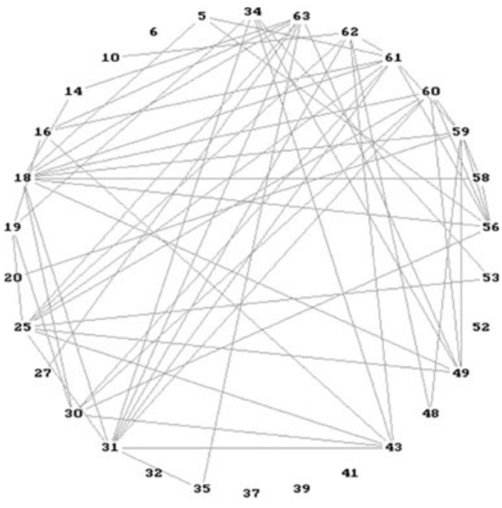


Figure 13.4 Knowledge sharing in Cortina

In a similar way:

$$CR_j = \frac{\sum_{i=1}^{n} (CT_{ji} * WC_i)}{\text{network size}}$$

CR_j = competence-based reputation of node 'j' (where $0 < RR_j < 1$)
CT_{ji} = effective competence-based trust relation of node 'j' with node 'i' (1 = activated; 0 = not activated)
WC_i = node's weight in terms of competence-based trust centrality (where $0 < WC_i < 1$)

The dataset was processed in two ways: first, three separate statistical analyses, one for each local network – Bormio, Cortina d'Ampezzo and Pila – were worked out; second, the entire dataset – as an aggregate of all data – was processed. Table 13.2 shows the statistical results of correlation analysis between relational reputation and centrality of a single actor inside the network.[5] The relational density value can be presented only in the case of the three effective networks, while it is a meaningless indicator for the analysis of the entire dataset.

These findings seem to confirm the following hypothesis:

(Hp1) There is a positive correlation between the degree of relational reputation and degree of centrality of a single actor inside the network.

The correlation between relational reputation and centrality for the entire dataset is quite significant (0.4116) and its control test confirms this result (p-value <0.0001).[6] It means that network-specific resources influence firm variety in terms of social dynamics. Figure 13.5 shows the QQ plot related to this result and highlights a positive trend especially in the central part of the graph.

Table 13.2 Analysis of the statistical correlation between relational reputation and centrality of a single actor inside the network

Dataset	Correlation between 'RR' and 'centrality'	p-value ($\alpha = 0.10$)	Network density (in terms of RR)	Network density (in terms of centrality)
Entire dataset	0.4116	<0.0001	/	/
Bormio	0.4229	0.0005	0.095	0.193
Cortina	0.5017	0.0047	0.062	0.143
Pila	0.3905	0.0971	0.163	0.322

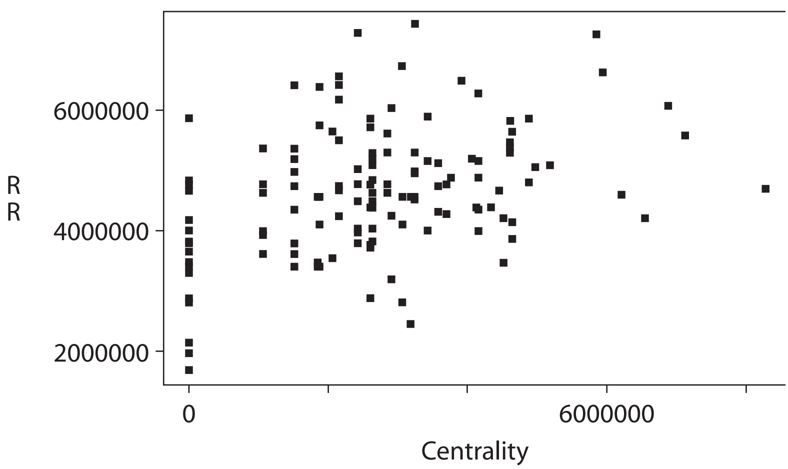


Figure 13.5 Correlation between relational reputation and centrality (QQ plot)

The sub-results with reference to the three contexts point to the same conclusion: relational reputation is closely correlated with the central position in the network. Table 13.2 shows also the degree of network density, calculated by using the adjacency matrix of knowledge sharing. When comparing the correlation values with these densities, a low network density seems to increase the positive association between the considered variables. Probably, a non-homogeneous distribution of the knowledge sharing flows – that is, the existence of many relevant sub-networks – amplifies the concentration of reputation and, therefore, the firm variety in terms of social assets. Maybe, when network density is high, and, hence, there is frequent and intense knowledge sharing, the role of relational reputation stays in the background. On the contrary, in a network characterized by a weaker exchange of knowledge – low density – relational reputation plays an important role and stands out in terms of polarization of knowledge flows and develops significant lock-ins.

The second research hypothesis analyses the same problem in terms of competence-based reputation and the findings support it as well. Structural indicators and the QQ plot, shown in Table 13.3 and Figure 13.6, lead to similar conclusions, but with some important differences.

First of all, the absolute levels of correlation between competence-based reputation

Table 13.3 Analysis of the statistical correlation between competence-based reputation and centrality of a single actor inside the network

Dataset	Correlation between 'CR' and 'centrality'	p-value (α = 0.10)	Network density (in terms of CR)	Network density (in terms of centrality)
Entire dataset	0.3903	<0.0001	/	/
Bormio	0.3380	0.0016	0.127	0.193
Cortina	0.5045	0.0045	0.037	0.143
Pila	0.3372	0.0911	0.069	0.322

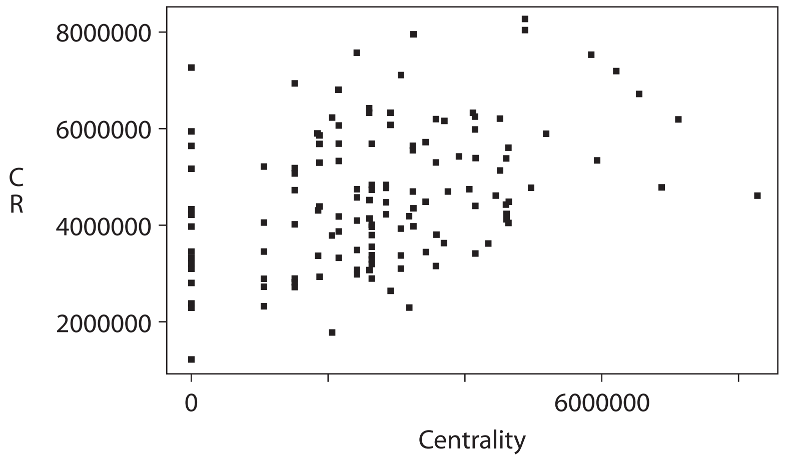


Figure 13.6 Correlation between competence-based reputation and centrality (QQ plot)

and centrality are again interesting, in terms of high value, but are lower than in the previous hypothesis. Social dynamics based on relational elements are probably more inertial and path dependent compared to competence-based trust/reputation. It empirically highlights the relevance of social dynamics in evolutionary economic geography. A second observation arising from these findings is that, in the case of competence-based reputation, the correlation with centrality is stronger when network density is low, even if the effect is less clear compared to the analysis of relational reputation. It may be argued that the advantage given by reputation, both relational or competence-based, is a relative advantage that depends on the inter-organizational context and district features. According to our empirical findings, good reputation is a strong competitive advantage especially when there is a weak exchange of knowledge, while there is an average advantage when there is an intense web of social relation and opinion sharing.

5. Conclusions and implications for local policies

Evolutionary economic geography explains the spatial evolution and variety of firms, local networks and clusters (Boschma and Frenken, 2006). This survey adds that social variety – in terms of reputation – within districts is a critical variable in an evolutionary perspective that is strictly related with knowledge sharing dynamics. In this framework, social dynamics plays a crucial role in the process of diversification of the firms in an evolutionary perspective. According to our empirical findings, social paths seem more inertial when compared to knowledge development trajectories.

In an inter-organizational structure without formal hierarchy – such as a local network among independent agents – reputation and trust may represent a crucial issue in order to explain firm evolution. We have found that good reputation can over time polarize knowledge flows towards a firm and influence its evolution pattern, thus improving its strategic positioning inside the network structure (in terms of centrality). It is an agreed opinion that reputation is a strategic asset (Dentchev and Heene, 2003) and that it can be easier for firms with a positive reputation to build and maintain networks based on reciprocal trust (Noorderhaven et al., 2002).

These connections give a sustainable competitive advantage to the actors in the local network because it is unique and impossible to imitate and because they shape the internal accumulation of (intangible) resources (Leana and Van Buren, 1999). According to our results, it is important to adopt a network structure perspective in order to develop local policies. The aim is to create a 'learning local environment' that supports evolution and diversification of (inter-)organizational routines based also on the appraisal of social heterogeneity (Asheim and Cooke, 1999; Boschma and Lambooy, 1999). These propositions lead to a rethink of local development policies by enhancing both firm- and network-specific resources, by considering both business and social variety. A recommendation arising from our research is that managers, entrepreneurs and local institution administrators should improve their ability to know, to read and to interpret social networks and their trajectories, which means recognizing firm differences too.

Notes

* This chapter is the outcome of a strict collaboration between the three authors. However, A. Zucchella wrote sections 1 and 2, S. Denicolai wrote sections 3 and 4, and G. Cioccarelli wrote section 5.

1. For instance, see Chapters 7, 11 and 15.
2. Founded in 1969, 'Mori' is a growing market and public opinion research agency. See http://www.mori.com/reputation/.
3. Because of the redemption rates, the study of centrality degrees refers to the related sub-networks within the tourism cluster as a whole.
4. The questionnaire reported a set of key definitions about the critical notions used, in order to guarantee a shared comprehension among all respondents.
5. The data processing was developed with 'SAS Enterprise Miner'. The *p*-value refers to the correlation index.
6. According to the research model in Figure 13.1, the aim of the datamining process was to check only the bidirectional correlation between trust – network-specific resources – and reputation heterogeneity – firm-specific resources – instead of proposing a regression model for reputation estimation, which implies the identification of dependent and independent variables.

References

Almeida, P. and B. Kogut (1999), 'Localisation of knowledge and the mobility of engineers in regional networks', *Management Science*, **45** (7), 905–17.

Aringhieri, R., E. Damiani, S. De Capitani Di Vimercati, S. Paraboschi and P. Samarati (2006), 'Fuzzy techniques for trust and reputation management in anonymous peer-to-peer systems', *Journal of the American Society for Information Science and Technology*, **57** (4), 528–37.

Asheim, B.T. and P. Cooke (1999), 'Local learning and interactive innovation networks in a global economy', in Edward J. Malecki and Päivi Oinas (eds), *Making Connections: Technological Learning and Regional Economic Change*, Aldershot: Ashgate.

Barber, B. (1983), *The Logic and Limits of Trust*, New Brunswick, NJ: Rutgers University Press.

Bathelt, H. and J. Glückler (2003), 'Towards a relational economic geography', *Journal of Economic Geography*, **3** (2), 117–44.

Becattini, G. (1990), 'The Marshallian industrial district as a socio-economic notion', in F. Pyke, G. Becattini and W. Sengenberger (eds), *Industrial Districts and Local Economic Regeneration*, Geneva: International Institute for Labor Studies, pp. 37–51.

Bellandi, M. (1989), 'The industrial district in Marshall', in E. Goodman and J. Bamford (eds), *Small Firms and Industrial Districts in Italy*, London: Routledge, pp. 136–52.

Boschma, R.A. (2004), 'Competitiveness of regions from an evolutionary perspective', *Regional Studies*, **38** (9), 1001–14.

Boschma, R.A. and K. Frenken (2006), 'Why is economic geography not an evolutionary science? Towards an evolutionary economic geography', *Journal of Economic Geography*, **6** (3), 273–302.

Boschma, R.A. and R.C. Kloosterman (eds) (2005), *Learning from Clusters. A Critical Assessment from an Economic-Geographical Perspective*, Dordrecht: Springer-Verlag.

Boschma, R.A. and J.G. Lambooy (1999), 'Evolutionary economics and economic geography', *Journal of Evolutionary Economics*, **9** (4), 411–29.

Boschma, R.A. and A.L.J. Ter Wal (2007), 'Knowledge networks and innovative performance in an industrial district: the case of a footwear district in the south of Italy', *Industry and Innovation*, **14** (2), 177–99.

Brady, K.O. (2002), 'Corporate reputation and sustainability: creating value by unifying two disciplines', *Ethical Corporation Magazine*, September.

Breschi, S. and F. Lissoni (2001) 'Knowledge spillovers and local innovation systems: a critical survey', *Industrial and Corporate Change*, **10** (4), 975–1005.

Brock-Smith, J. (1997), 'Selling alliances, issues and insights', *Industrial Marketing Management*, **26**, 149–61.

Brown, J.S. and P. Duguid (1991), 'Organizational learning and communities of practice: toward a unified view of working, learning, and innovation', *Organization Science*, **2** (1), 40–57.

Bunnell, T.G. and N.M. Coe (2001), 'Spaces and scales of innovation', *Progress in Human Geography*, **25** (4), 569–89.

Burt, R.S. (1992), *Structural Holes: The Social Structure of Competition*, Cambridge, MA: Harvard University Press

Cao, Q. and M.J. Schniederjans (2006), 'Agent-mediated architecture for reputation-based electronic tourism systems: a neural network approach', *Information and Management*, **43** (5), 598–606.

Castells, M. (1996), *The Rise of the Network Society*, Oxford: Blackwell.

Cioccarelli, G. (2003), *Turismo Alpino e Innovazione* [Alphine Tourism and Innovation], Milano: Giuffrè.

Cohen, W.M. and D.A. Levinthal (1990), 'Absorptive capacity: a new perspective on learning and innovation', *Administrative Science Quarterly*, **35**, 128–52.

Cooke, P. (1998), 'Regional innovation system. An evolutionary approach', in H. Braczyk, P. Cooke and R. Heidenreich (eds), *Regional Innovation Systems*, London: UCL Press.

Cooke, P. and R. Huggins (2002), 'High-technology clustering in Cambridge (UK)', in A. Amin, S. Goglio and F. Sforzi (eds), *The Institutions of Local Development*, London: Ashgate.

Cummings, L.L. and P. Bromiley (1996), 'The organizational trust inventory (OTI): development and validation', in R.K. Kramer and T.R. Tyler (eds), *Trust in Organizations: Frontiers of Theory and Research*, London: Sage, pp. 302–30.

Daft, R.L. (2001), *Organization Theory and Design*, 7th edn, Cincinnati, OH: South-Western College Publishing.

Dahl, M.S. and C.O.R. Pedersen (2005), 'Social networks in the R&D process: the case of the wireless communication industry around Aalborg, Denmark', *Journal of Engineering and Technology Management*, **22** (1–2), 75–92.

Dei Ottati, G. (1994), 'Trust, interlinking transactions and credit in the industrial district', *Cambridge Journal of Economics*, **18**, 529–46.

Denicolai, S. and A. Zucchella (2005), 'Territorial marketing and strategic management for local development: which connections and contradictions? Some reflections on Italian experiences', in H. Szulce and M. Florek (eds), *Marketing Terytorialny* [Territorial Marketing], Poznan: WAEP.

Denicolai, S. and A. Zucchella (2006), *Strategic and Organizational Analysis for Local Development: The Development Migration Path*, milan: Franco Angeli.

Dentchev, N.A. and A. Heene (2003), 'Reputation management: sending the right signal to the right stakeholder', Working Paper, Gant University.

Dobing, B. (1993), 'Building trust in user–analyst relationships', Doctoral Dissertation, Informal and Decision Sciences Department, University of Minnesota, Minneapolis.

Donet, P.M. and J.P. Cannon (1997), 'An examination of the nature of trust in buyer–seller relationships', *Journal of Marketing*, **61**, 35–51.

Elangovan, A.R. and D.L. Shapiro (1998), 'Between trust and control, developing confidence in partner cooperation in alliances', *Academy of Management Review*, **23** (3), 547–66.

Ely, J. and J. Valimaki (2003), 'Bad reputation', NAJ Economics, Northwestern University, Working Paper, Charles University, Prague.

Engelmann, D. and U. Fischbacher (2003), 'Indirect reciprocity and strategic building in an experimental helping game', Working Paper, Egham.

Essletzbichler, J. and D.L. Rigby (2007), Exploring evolutionary economic geographies', *Journal of Economic Geography*, **7** (5), 635–49.

Fombrun, C.J. (2001), 'Corporate reputation as economic asset', in M.A. Hitt, E.R. Freeman and J.S. Harrison (eds), *The Blackwell Handbook of Strategic Management*, Oxford: Blackwell.

Fombrun, C.J. and M. Shanley (1990), 'What's in a name? Reputation building and corporate strategy', *Academy of Management Journal*, **33** (2), 233–58.

Foss, N.J. (1999), 'Networks, capabilities, and competitive advantage', *Scandinavian Journal of Management*, **15**, 1–15.

Gambetta, D. (1998), *Can We Trust Trust? Making and Breaking Cooperative Relations*, New York: Blackwell.

Garofoli, G. (1991), *Modelli locali di sviluppo* [Local Patterns of Development], Milano: Franco Angeli.

Gemunden, H.G., T. Ritter and P. Heydebeck (1996), 'Network configuration and innovation success: an empirical analysis in German high-tech industries', *International Journal of Research in Marketing*, **13**, 449–62.

George, V.P. and T.J. Allen (1989), 'Net graphs: a graphic representation of adjacency matrices as a tool for network analysis', MIT, International Center for Research on the Management of Technology, Working Paper.

Glückler, J. (2007), 'Economic geography and the evolution of networks', *Papers in Evolutionary Economic Geography*, **7** (4), Utrecht University, Netherlands.

Granovetter, M. (1985), 'Economic action and social structure: the problem of embeddedness', *American Journal of Sociology*, **91**, 481–510.

Gulati, R. (1995), 'Does familiarity breed trust? The implications of repeated ties for contractual choice in alliances', *Academy of Management Journal*, **30** (1), 471–85.

Gulati, R. (1999), 'Network location and learning: the influence of network resources and firm capabilities on alliance formation', *Strategic Management Journal*, **20** (5), 397–420.

Gulati, R. and M. Gargiulo (1999), 'Where do interorganizational networks come from?', in *American Journal of Sociology*, **104** (5), 1439–93.

Hansen, M.T. (1999), 'The search-transfer problem: the role of weak ties in sharing knowledge across organizational subunits', *Administrative Science Quarterly*, **44**, 82–111.

Kenis, P. and D. Knoke (2002), 'How organizational field networks shape interorganizational tie-formation rates', *Academy of Management Review*, **27** (2), 275–93.

Kilduff, M. and W. Tsai (2003), *Social Networks and Organizations*, London: Sage.

Knoke, D. and J.H. Kuklinski (1982), *Network Analysis*, Newbury Park, CA: Sage.

Kogut, B. and U. Zander (1993), 'Knowledge of the firm and the evolutionary theory of the MNC', *Journal of International Business Studies*, **24** (4), 625–45.

Köszegi, S. (2002), 'Do we have enough trust in trust?', Process of International Negotiation, IIASA.

Krackhardt, D. (1990), 'Assessing the political landscape: structure, cognition and power in organization', *Administrative Science Quarterly*, **35**, 342–69.

Kreps, D.M. and R. Wilson (1980), 'Reputation and imperfect information', *Journal of Economic Theory*, **27**, 253–79.

Leana, C.R. and H. Van Buren III (1999), 'Organizational social capital and employment practices', *Academy of Management Review*, **24**, 538–55.

Maggioni, M.A. and T.E. Uberti (2005), 'International networks of knowledge flows: an econometric analysis', Papers in Economics and Evolution, Max Planck Institute, Jena.

McAllister, D.J. (1995), 'Affect and cognition-based trust as foundations for interpersonal cooperation in organizations', *Academy of Management Journal*, **38** (1), 24–59.

McKnight, D., L. Cumming and N. Chervany (1995), 'Trust formation in new organization relationship', Working Paper, University of Minnesota.

Malmberg, A. and P. Maskell (2002), 'The elusive concept of localization economies. Towards a knowledge-based theory of spatial clustering', *Environment and Planning A*, **34** (3), 429–49.

Marshall, A. (1920), *Principles of Economics*, 8th edn, London: Macmillan.

Merry, S.E. (1984), 'Rethinking gossip and scandal', in D. Black (ed.), *Towards a General Theory of Social Control*, Vol.1, New York: Academic Press, Inc.

Morgan, K. (1997), 'The learning region: institutions, innovation and regional renewal', *Regional Studies*, **31**, 491–503.

Nahapiet, J. and S. Goshal (1998), 'Social capital, intellectual capital and the organizational advantage', *Academy of Management Review*, **23**, 242–66.

Nelson, R. (1993), *National Innovation Systems: A Comparative Analysis*, Oxford: Oxford University Press.

Nelson, R.R. and S.G. Winter (1982), *An Evolutionary Theory of Economic Change*, Cambridge, MA: Belknap Press of Harvard University Press.
Nieminen, J. (1974), 'On centrality in a graph', *Scandinavian Journal of Psychology*, **15**, 322–36.
Noorderhaven, N.G., C.I. Koen and S. Beugelsdijk (2002), 'Organizational culture and network embeddedness', Discussion Paper No. 2002/91, Tilburg University, Netherlands.
Nooteboom, B. (2002), *Trust: Forms, Foundations, Functions, Failures and Figures*, Cheltenham, UK and Northampton, MA, USA: Edward Elgar.
Nooteboom, B. (2004), *Interfirm Collaboration, Learning and Networks*, London: Routledge.
Nooteboom, B. and V.A. Gilsing (2004), 'Density and strength of ties in innovation networks: a competence and governance view', Report Series, Erasmus Research Institute of Management.
Nooteboom, B., H.E. Berger and G. Noorderhaven (1995), 'Sources, measurement and effect of trust in the governance of buyer–supplier relations', Working Paper, School of Management and Organization, Groningen University.
Perrone V. (2001), 'La collaborazione interorganizzativa in un distretto turistico: la rete dell'ospitalità' [Interorganizational collaboration in a tourist district: the network of hospitality], in S. Salvemini (ed.), *Artwork and Network*, Milano: Egea.
Powell, W. (1990), 'Neither market nor hierarchy', *Research in Organizational Behavior*, **12**, 295–336.
Putnam, R.D. (1993), *Making Democracy Work: Civic Tradition in Modern Italy*, Princeton, NJ: Princeton University Press.
Ring, P.S.E. and A.H. Van der Ven (1994), 'Developmental processes of cooperative interorganizational relationship', *Academy of Management Review*, **19** (1), 90–118.
Ritter, T. (1999), 'The networking company, antecedents for coping with relationships and networks effectively', *Industrial Marketing Management*, **28**, 467–79.
Roberts, P. and G.R. Dowling (2000), 'Reputation and sustained superior financial performance', *Academy of Management Proceedings*, 1–6.
Rousseau, D.M., S.B.E. Sitkin, R.S. Burt and C. Camerer (1998), 'Not so different after all: a cross-discipline view of trust', *Academy of Management Review*, **23** (3), 393–404.
Saxenian, A. (1994), *Regional Advantage: Culture and Competition in Silicon Valley and Route 128*, Cambridge, MA: Harvard University Press.
Saxton, T. (1997), 'The effects of partner and relationship characteristics on alliance outcomes', *Academy of Management Journal*, **40**, 443–61.
Sforzi, F. (1990), 'The quantitative importance of Marshallian district in the Italian economy', in F. Pyke, B. Becattini and W. Sengenberger (eds), *Industrial Districts and Inter-Firm Cooperation in Italy*, Geneva: International Institute for Labour Studies.
Shapiro, C. (1983), 'Premiums for high quality products as returns to reputations', *The Quarterly Journal of Economics*, **98** (4), 659–79.
Shapiro, S.P. (1987), 'The social control of impersonal trust', *American Journal of Sociology*, **93**, 623–58.
Simonin, B.L. (1997), 'The importance of collaborative know-how: an empirical test of the learning organization', *Academy of Management Journal*, **40**, 1150–74.
Storper, M. and R. Walker (1989), *The Capitalist Imperative. Territory, Technology and Industrial Growth*, New York: Basil Blackwell.
Sutter, M. and M.G. Kocher (2004), 'Age and the development of trust and reciprocity', Papers on Strategic Interaction 2004–01, Max Planck Institute of Economics, Strategic Interaction Group.
Torre, A. and A. Rallet (2005), 'Proximity and localization', *Regional Studies*, **39** (1), 47–60.
Uzzi, B. (1996), 'The sources and consequences of embeddedness for the economic performance of organizations: the network effect', *American Sociological Review*, **61**, 674–98.
Van Haverbeke, W. and N.G. Noorderhaven (2001), 'Competition between alliance blocks: the case of the RISC-microprocessor technology', *Organization Studies*, **22**, 1–30.
Winter, S.G. (2002), 'Understanding dynamic capabilities', Working Paper of Reginald H. Jones Center, The Wharton School, University of Pennsylvania, WP2002–05.
Xiong, L. and L. Liu (2003), *A Reputation-Based Model for Peer-to-Peer E-commerce Communities*, Atlanta: College of Computing, Georgia Institute of Technology.
Zucchella, A. and A. Ganzaroli (2003), 'Creating trust between local and global networks. Evidence from Italian industrial districts', in Proceedings of 2nd Workshop on Trust Between and Within Organisations, Amsterdam, 23 and 24 October 2003, EIASM and Free University of Amsterdam.
Zucchella, A., M.E. Maccarini and P. Scabini (2004), 'Co-location and business-specific issues in the internationalisation of SMEs. Their impact on the growth and on the international performance of born global firms', *Proceedings of the University of Vaasa, Finland, Reports*, **111**, 131–49.
Zucker, L.G. (1986), 'Production of trust: institutional sources of economic structure', in B. Staw and L. Cummings (eds), *Research in Organizational Behavior*, 8, Greenwich, CT: JAI Press.

14 The evolution of a strategic alliance network: exploring the case of stock photography

Johannes Glückler

1. Introduction

Over recent years, increasing effort has been made to conceptualize an evolutionary perspective within economic geography (Boschma and Frenken, 2006; Boschma and Lambooy, 1999). Evolutionary approaches have theorized a variety of conceptions, among others, the notion of path dependence in regional development (Martin and Sunley, 2006), the role of technological and sectoral variety in regional innovation (Boschma and Iammarino, 2009; Rigby and Essletzbichler, 2006) and the idea of co-evolution in regional development (Schamp, Chapter 20, this book). Explicitly or implicitly, many geographical accounts of the economy rely on the notion of networks to represent and analyse economic processes and modes of organization. A dynamic perspective of network evolution, however, has experienced only little attention until recently (Baum et al., 2003; Kilduff and Tsai, 2003; McPherson et al., 2001). The common research strategies to focus on the temporal dimension of networks have been twofold. On the one hand, contingency models of networks have analysed the appropriateness of certain network forms for different environmental conditions. Empirical research demonstrates, for example, that when firms operate under conditions of exploration, that is, high uncertainty and high innovation rates, weak ties and loose coupling are more beneficial than strong ties. In turn, when firms operate under conditions of exploitation, that is, stable market conditions and mature products, strong ties are more advantageous (Rowley et al., 2000). On the other hand, sequential models of non-linear cyclical change have been applied to the evolution of firms (Hite and Hesterly, 2001; Lavie, 2004) and regional clusters (Staber, 1997). Within the framework of organizational ecology, Staber (1997) found that in the beginning of a textile cluster, dense networks with strong ties between firms were beneficial in order to circulate knowledge and to ally for joint lobbying. In later stages of cluster development, however, when an increasing number of firms competed for the same resources, that is, labour, expertise, capital and so on, weak ties became more important (Staber, 2001). Though these models recognize the ephemeral nature of network structure, they are relatively silent on the endogenous mechanisms that drive network evolution because they always use external environmental changes to explain responses in network development.

This chapter aims at exploring and connecting network evolution with the wider project of an evolutionary economic geography (Glückler, 2007). Concretely, this chapter pursues two questions. First, are there endogenous network dynamics, that is, mechanisms within network structure that drive the dynamics of inter-firm networks and help to explain future network structure? Second, is geographical proximity associated with growth patterns of these inter-firm networks? The empirical challenge in evolutionary network research is to find appropriate longitudinal structural data on sectoral or regional economic activity. This chapter focuses on the inter-firm sales partnerships in the organizational field of stock photography and uses a survey on German stock photo agencies for the period between 1989 and 2005. Stock photography is an interesting sectoral case since it has been undergoing major technological, institutional and organizational change in a very short time (this shift is discussed in more detail below). The empirical analysis looks at the recent emergence of sales alliances between picture agencies as the unit of analysis. Strategic alliances are an organizational strategy to access and share resources, costs and profits with partner organizations. Especially under conditions of increased competition and unstable market conditions, they render different advantages to individual organizations, such as firm survival, performance, imitation and innovation (Brass et al., 2004; Knoke, 2001). A sales partnership, in particular, is a contractual licence agreement, where one firm grants the usage right either for an intellectual property-right protected content to another firm for a specified time and use in return for a royalty payment (Knoke, 2001). Though stock photography has a long tradition in international sales cooperation (Wilkinson, 1997), the alliance behaviour has changed dramatically over the last ten years. It has changed in quantitative terms because more and more picture agencies seek sales partners; and it has experienced a qualitative change because picture agencies no longer ally only internationally but they have also started partnering with agencies in their domestic market. Section 2 discusses some elements of an emerging theory of network evolution and develops research hypotheses. Section 3 introduces the business context and the fundamental technological and organizational changes occurring in the market for still images in recent history. Section 4 documents the methodology and section 5 presents the findings of the empirical analysis before the chapter closes with some concluding remarks with respect to the emerging project of an evolutionary economic geography.

2. Network evolution and geography

The essential starting point for any theory about network evolution is the question of 'how do structural dimensions of an interorganizational communication network at Time 1 affect the interactions among member organizations – specifically, their formation of ties to other organizations – at Time 2?' (Kenis and Knoke, 2002, pp. 277–278). This question pays attention to inherent conditions of a network rather than external conditions that drive network change. The network trajectory (Kilduff and Tsai, 2003) is an appropriate concept in the analysis of network evolution, which combines the notions of evolution, network and geography. It describes a geographically and historically specific development path of a network in which the formation and dissolution of ties in earlier stages generate cumulative propensities for the formation and dissolution of ties in the future and in which the mechanisms of path-disruption and variation are endogenous (Glückler, 2007). This perspective explicitly moves beyond the dyadic analysis of single relations to the analysis of entire networks of relations. A theory of network evolution, thus, looks at the changes that new ties produce in the existing structure and, conversely, at the impact that the structure imposes on the formation of these next ties. In a recent, sophisticated study Powell et al. (2005) explicitly tested a set of alternative hypotheses on the emergence of new relationships in an expanding network. Following on the network literature, they focused on a set of hypotheses that they assessed as potential mechanisms for network evolution: homophily, preferential attachment and multiconnectivity. The following sections briefly review these concepts before geography is explored as a fourth potential condition of network evolution.

Homophily vs. diversity between firms

According to homophily theory, firms with similar attributes have greater ease in developing characteristic-based trust (Zucker, 1986) and are therefore more likely to engage in trust-demanding activities. The similarity between firms implies a stronger overlap of interests, strategies and interpretative frameworks to deal with information on markets and products. Hence, networks are expected to expand by a process in which new partners are chosen for similarity or homophily (Rowley et al., 2005; Sorenson, 2003; McPherson et al., 2001). The homophily between firms may be measured as the similarity in firm size, age and form as well as similarities in terms of strategy and products. Alternatively, however, alliances may also be hypothesized to follow a pattern of diversity. Firms may choose to collaborate exactly because of different, complementary resources and competencies (Rowley et al., 2005) in order to yield synergies from collaboration. Therefore, the dissimilarity with respect to strategy, resource endowment and so on may be also be a driving element in the formation of strategic alliances.

Preferential attachment and multiconnectivity

Apart from firm-specific characteristics that may be compared and assessed for pair-wise homophily, Powell et al. (2005) also propose structural effects on future tie formation that would produce endogenous dynamics. The preferential attachment hypothesis expects firms with many ties at one point in time to be more likely to receive new ties in the future than those with fewer ties. There is obviously an accumulative advantage for well positioned actors on the one side and a liability of disconnectedness for peripheral actors on the other, as the authors found in earlier research (Powell et al., 1996). Alternatively, the multiconnectivity hypothesis expects networks to expand through a process in which firms seek diversity of relations and form multiple independent paths. By moving in different communities and connecting to heterogeneous groups, the firm enhances the multiple reachability of partners. Empirically, however, Powell et al. largely disprove homophily and popularity by analysing extensively the evolution of the strategic alliance network in US biotechnology over a 12-year period between 1988 and 1999. Instead, they found strong support for the multiconnectivity hypothesis, because new alliances were more likely to form between those firms that were more multiply or more diversely linked to each other at a previous stage of network growth. This finding suggests that networks entail a logic of multiconnectivity, where new linkages reinforce diversity and multiple cohesion (Powell et al., 2005). Another strand of network research supports these conclusions from a different perspective. While the former approach emphasizes the mechanisms of change through the appearance of new ties in an incumbent network, Walker et al. (1997) emphasize path-dependence and persistence in network structure through its evolution over time. In their empirical analysis of the alliance network in biotechnology they found that early partner choice at the beginning had a significant impact on future cooperation. A year-to-year comparison of structural equivalence for each pair of actors showed that if two firms were connected to the same other companies, that is, they were structurally equivalent, they most likely remained so throughout the subsequent expansion of the network. This finding leads Walker et al. (1997) to argue for path dependence in the network trajectory.

Geographical proximity

Interestingly, a lot of the economic evidence on path dependence and lock-in has actually been produced and exemplified in the context of geographic clusters (Martin and Sunley, 2006). Geographical propinquity seems to be an important constraint on network formation. Sometimes this effect is abbreviated in unfortunate terms of 'spatial causation'. Space, however, is not a necessary cause of economic interaction. Instead, its role is mediated through at least two underlying social technologies: communication technology (Storper and Venables, 2004) and transport technology (Marquis, 2003). Breschi et al. (Chapter 16, this book) demonstrate nicely that the effect of geographical proximity on innovation is often mediated through inert mobility of inventors. They tend to stay in a region over long periods of time but move between companies in the same region to produce a pattern of locally interrelated innovation networks. Only with respect to the actors' communication preferences and mobility opportunities may the contingent relation between physical space and economic interaction be established (Glückler, 2007). In other words: the constraints of proximity only rule if face-to-face is the only mode of communication and if travel is prohibitive. In any other case, especially in a world of high mobility of people, commodities and information, proximity is contingent on the underlying social technologies. Owen-Smith and Powell (2004) distinguish two such communication technologies to explain why knowledge-intensive businesses such as biotechnology tend to favour co-located patterns of inter-firms alliances: one is the *conduit* as the distribution of private information from a sender to a number of limited receivers, and the other is the *channel* as the potential spill-over of information into the entire set of connected actors in a network under conditions of spatial proximity (Owen-Smith and Powell, 2004). In fact, the more complex the knowledge in a particular industry, the more do industries agglomerate. This is mainly because the distribution of complex knowledge entails more problems of accurate transmission and interpretation such that spatial proximity eases transfer and locks out remote actors from the knowledge flow (Fleming and Sorenson, 2001; Sorenson, 2005). Furthermore, Powell et al. (2005) found unequivocal evidence for a strong geographical bias on strategic alliances in the organizational field of biotechnology. New ties as well as repeat ties were more likely when two firms were co-located. These studies univocally suggest the importance of geography to the organization and dynamics of networks.

Geographical proximity is not only mediated through communication technologies at a given moment in time but may also be a result of its historical legacy. Marquis's (2003) work illustrates just how geographical proximity may continue to affect the formation of network relations even though communication and travel technologies have changed dramatically. He demonstrates that those networks of interlocking directorships in US cities that had been established prior to the advent of air travel technology continued to be significantly more locally bound than networks in younger cities. Despite the availability of modern travel technologies in all cities today, even new corporate board positions were filled with local directors. This persistence of geographical network structure in older communities supports imprinting theory (Stinchcombe, 1965): organizations adopt organizational characteristics in response to the environmental conditions during their period of foundation. Marquis's study legitimizes the possibility that geographical proximity endures in network practice even though the communication and travel technologies have changed and now allow for different patterns of spatial interaction. These arguments are challenging with respect to sales alliances in the case of stock photography because digital technology has enabled picture agencies to operate nearly independent of physical constraints and to exchange pictures with other agencies in any location. Will geography still matter for the creation of new alliances in the network, although the business processes are almost fully digitized and virtualized?

3. A case from the cultural economy: the evolution of stock photography

Photography is an essential element of visual information and communication at all levels of society. We encounter still images in press, television, the internet, books, exhibitions, corporate communication, and advertisement campaigns. Wherever we find a photograph, there is a supply chain behind the placement and there are basically two ways for a photo to reach the market. One is to assign a photographer and have photos produced in photo-shootings, the other is to buy a licence for the use of an existing photograph from a stock picture agency. Licensing a stock photo has a number of advantages. In contrast to a photo production, the customer avoids the production risk, that is, she sees what she gets, and saves the production cost, which might add up to considerable sums for location, equipment, models, stylists and the photographer crew. The customer only pays a royalty to the agency for using a photo, which might be just a small percentage of the production cost. Typically, it is the picture editor in a publishing house or the art buyer in an advertising agency who takes the decision whether to assign a photographer for production or to buy a photo 'from the shelves'. In the USA, picture editors and art buyers spend about double the amount of money on stock photography than on photo-shootings (Sachs, 2003) and use about four times more stock photos than their own productions (Kjemtrup, 1997). Today, there are around 1000 picture agencies in Europe offering photographs and illustrations from about 21,000 photographers and generating revenues of roughly 1 billion euros (CEPIC, 2001).

The historical evolution of stock photography might well be interpreted as a continuous response to technological innovation. Ever since its invention by Louis Daguerre in Paris and W.H. Fox Talbot in London, photography has been shaped and developed through the ongoing advance in production and transmission technologies: Talbot was the first to develop the principle of the negative pattern, an invention that laid the ground for repeated duplication in 1840 (Frizot, 1998); and Arthur Korn's invention of the *Telefotokopie* in 1902, for instance, built the technological base for long-distance circulation and distribution of photographs within a short time and thus made photography available as a visual medium in the established news press at the beginning of the twentieth century (Albert and Feyel, 1998). The emergence of picture agencies as intermediaries between photographers and publishing houses was a consequence of the rapid growth of photography in the 1920s. Toward the end of the 1930s a significant number of firms specialized in archiving and marketing photography on commission of photographers. After this historical departure, the organizational field of stock photography experienced some fundamental changes. Three major organizational and technological shifts are apparent in this context (see Glückler, 2005 for an extended elaboration).

From press to publicity: expansion into commercial photography

Stock photography had long been editorial, that is, pictures for reportage, documentation, and news, whereas advertising agencies would usually produce their visual content through assignment photography with commercial photographers. The 1970s brought about a major change. Given the economic downturn and the oil crisis in 1973, the advertising industry suffered serious budget constraints and sought new image sources at lower cost. At the same time, the 1978 US Copyright Law, which asserted that an image was the photographer's rather than the client's property, provided an institutional guarantee for photographers to use stock photography as an additional business opportunity (Frosh, 2001). *The Image Bank* pioneered modern stock photography by venturing from the editorial into the commercial market in 1974. In order to compete with assignment photography, they managed to improve picture quality, attract commercial photographers, and attract the attention of advertising agencies to buy stock at a lower cost than that of assigning photo productions. The company pursued an aggressive marketing strategy, introducing picture catalogues as an instrument to gain visibility and create visual style (Frosh, 2001). Today, picture catalogues are an established marketing tool. In Europe, half of all agencies publish catalogues on an annual basis (CEPIC, 2001). In addition, the British pioneer Tony Stone developed a sophisticated licensing system to maximize the profit per picture such that the agency succeeded in selling the same picture over 300 times in just one year, generating revenues of £30,000 with just one photograph (Rich, 1994).

From light-box to screen: digitization and online archives

In the early 1990s the market experienced enormous growth rates of up to 25 per cent per year and enthusiastic expectations accompanied the advance of digital technology. However, still in 1995 no single company had launched an online archive, and a survey in Britain at the time revealed that many of the big agencies had not yet taken clear decisions when to go online. Some business experts were even quite averse to the notion of selling pictures via the internet and saw more threat than opportunity: 'I can't understand why any picture library can consider going on-line' (Sal Shuel, BAPLA, quoted in Goddard 1995). While the incumbent business was somehow reluctant to embrace the advent of digital technology, very powerful and resourceful investors entered the market with high expectations. The strategic opportunity of digitization was based on the fact that more images could be offered and distributed to more customers at essentially less cost. British oil magnate Mark Getty and Microsoft founder Bill Gates recognized this opportunity and made large investments to buy up basically all established and significant players in the market. Today, Getty Images (founded 1995) and Corbis (founded 1989) are by far the largest firms in the global market place for still images. In contrast to their incumbent competitors, both firms drove technological change with intensive commitment. Getty Images were the first big agency to launch an e-commerce platform on their online archive in 1998. While they gained 14 per cent of their revenues in the first year, they increased online sales to over 90 per cent in only five years.

From licence to property: royalty-free images

Online archives and e-commerce platforms have not been the last innovation, though. Decreasing distribution cost through digital technology has made another source of friction more obvious: the legal constraints of intellectual property rights. When customers buy a reproduction licence, the royalty is tailored to the intended use and depends on the size and placement of the photography, the medium (commercial or editorial), the print run, and eventually on exclusivity. If the customer wanted to extend the use of a photograph to, for example, brochures, homepage, flyers, posters and so on she would have to buy a new reproduction licence and pay an additional royalty. Managing the property and reproduction rights implies transaction cost and limits the flexibility of the customer. As a consequence, a new licence model called 'royalty-free' was successfully established in stock photography. When customers buy a royalty-free (RF) photo they pay a fixed price and acquire unlimited usage rights. The royalty-free licence model thus converts photography into a full commodity. Mark Torrence had known this licence model in the music industry, when he founded PhotoDisc in 1991 to pioneer stock photography with RF-visual content. Servicing graphic and web designers in the beginning, PhotoDisc launched an e-commerce platform in 1995 and drove the company to be among the top ten of America's 500 fastest growing companies in 1997. Like many successful agencies, Getty Images acquired PhotoDisc in order to incorporate their technology and know-how and to go online just a year later in 1998. In 2000, the market share of RF-images was already assumed to reach 15 per cent in the US and 7 per cent in Europe. Today, Getty Images gain about one third of their revenues from royalty-free images and at higher margins per image than traditional rights-managed photography (Glückler, 2005).

In conclusion, the market for stock photography has experienced major changes, which have only been highlighted very briefly. Nonetheless, digital technology and royalty-free photos have dramatically shaken up the market environment, and redefined fundamental business parameters. As a consequence, the market now looks very different from what it looked like 15 years ago. It has shifted from a market of medium-sized companies to an oligopolistic market dominated by three players that easily take a 45 per cent share of the global revenues in this sector; it has moved from localized enduring personal relationships to a virtual business transaction; and it is transforming photography from a licensable content to a private commodity (Glückler, 2005). This shift has deep cutting implications for competitive strategy. The availability of e-commerce and internet search in online picture archives disrupts existing networks of personal relationships between agents and customers and turns a seller market into a buyer market. The global dominance of Getty Images and Corbis forces picture agencies into niche strategies, specialization and inter-firm alliances. Since market visibility and the scale of photo collections have become so centre stage, many agencies have begun to seek sales partnerships in order to forge alternative distribution channels or to collect additional material and reach scale.

The aim of the empirical analysis is to analyse the growth pattern of the stock photography field-net of inter-agency licence agreements in Germany. An organizational field-net is 'a particular pattern of both present and absent links among the entire set of organizational dyads occurring in a specified organizational field' (Kenis and Knoke, 2002). The term is used to explicitly focus on the dyadic relations and the overall structure of relations between every pair of players in a field.

4. Methodology

The empirical approach to test the hypotheses of homophily, geography, popularity bias, and multiconnectivity on the alliance behaviour of picture agencies made a network survey necessary because there are no secondary data available on alliance behaviour in the stock photography market. The survey method, however, limits the study of evolutionary processes in at least three ways: it cannot deliver long-term historical data; it will normally produce incomplete responses; and it produces a survival bias by missing out those firms and earlier alliances that were abandoned in the past. The network thus represents a growth path of surviving firms rather than the entire network evolution.

Variables

The questionnaire consisted of two parts. The first part collected attribute information on the individual picture agencies such as firm age, product specialization, firm size by number of employees, by image stock and by number of represented photographers as well as international sales and membership in the national business association. These attribute items were used to develop relational values of firm-to-firm similarity (homophily) and geographical propinquity between all pairs of agencies (see Table 14.1 for variable definitions). The calculation of relational similarities is based on measures of pair-wise differences or coincidence in the individual scores on an item. For each variable, the similarity values are then represented in a similarity matrix for all pairs of $n \times$

Table 14.1 Variable labels and definitions, dyad level

Label	Description
Dependent variable	
Sales alliances	Sales alliances between all pairs of agencies for each year, 1989–2005
Independent variables	
A. Homophily	
Firm age	Difference in years of foundation in Germany
Digital entry	Difference in years of launching an online picture archive
Size employment	Difference in number of employees
Size images	Difference in size of picture archive
Size photographers	Difference in number of photographers (suppliers)
Universality	Dummy, match for two firms as specialized (= 0) or universal agencies (= 1)
Specialization	Dummy, match for two firms according to the kind of specialization of picture portfolio
International sales	Difference in percentage of foreign annual sales
Lobby association	Dummy, match for two firms according to their membership in the national industry association BVPA
B. Geography	
Location	Dummy, match for two firms according to their location in the same metropolitan region
C. Popularity bias	
Indegree centrality	Difference in in-degree
Outdegree centrality	Difference in out-degree
D. Multiconnectivity	
Point connectivity	Number of different firms through which a pair of firms is connected
Geodesic count	Number of shortest paths between two agencies

n firms, where each cell value indicates the coincidence (e.g. association membership) or (dis)similarity (e.g. difference in firm age) of two firms on an item.

The second part asked firms to identify all sales partnerships with other agencies. The variable is represented in a binary adjacency matrix, where for each pair of agencies the value 1 indicates the existence of an alliance and 0 indicates its absence. Thus, the *dependent variable* is the network of sales alliances for trading rights-managed picture licences between picture agencies in Germany. Each relationship was further specified according to three characteristics: first, two types of licence model – rights-managed vs. royalty-free – were distinguished. Since the alliance behaviour in royalty-free imagery is quite different from cooperation in rights-managed photography, the study focused purely on the dominant business of rights-managed sales alliances. Second, the year of establishment of each licence agreement was assessed in order to produce network stages on a year-to-year basis. This temporal dimension of network formation builds the basis for the analysis of network dynamics over time and allows for testing the effects of homophily, geographical propinquity, popularity bias and multiconnectivity on later stages of network structure. Third, the direction of picture supply distinguishes between senders and receivers and thus identifies reciprocal and unidirectional linkages.

Finally, four other variables were generated from the alliance data. The dyadic differences between individual firms' in-degree and out-degree centralities were used to measure the concept of popularity bias. The greater the differences in in-degree (receiving relations) or out-degree (sending relations) centrality between a pair of agencies, the more likely should they be to have a relationship in the future. The concept of multiconnectivity was measured by two variables. The point connectivity of two unconnected (non-allying) firms is the number of firms that need to be deleted so that no indirect connection exists between them any more (Borgatti et al., 2002). Geodesic count produces a measure that counts the number of shortest network paths between each pair of agencies (Borgatti et al., 2002).

Survey population

There are three organizational types of picture agency. In the dominant case the agency is a commercial representative of the images provided by independent photographers. The photographer holds the intellectual property rights and commissions the agency to sell reproduction licences for a royalty. Apart from this dominant form, there are photographer cooperatives and archives. In contrast to agencies, the former maintain the intellectual property rights within the cooperative and the latter can market historical public domain images, where intellectual property rights have expired. Therefore cooperatives and archives may offer pictures at more competitive prices since they do not have to share their revenues with the legal originator. The population of picture agencies is defined as the number of commercial firms that license usage rights of still images under commission of independent photographers. This explicitly excludes non-commercial archives, small-scale sideline businesses and direct-marketing activities of photographers. Photographer cooperatives, however, are included in the analysis because they often license material also from independent photographers. For the year 2005, 201 agencies were identified through the industry association's member directory, own research and through snowball sampling in the course of data collection. Altogether, 75 agencies completed the questionnaire, where 70 had at least one sales alliance in rights-managed photography. The adjacency matrix of this alliance network between the set of 70 agencies represents the independent variable. Though the response rate of 37 per cent is not very high, the respondents represent the majority of the actual business in the market. Many medium and large players took part in the study and most non-respondents were micro businesses with often just one self-employed owner, and most of them unknown to the national business association. Therefore it seems justified to assume that the complete network would look quite similar to the one surveyed. The only alliances missing out in the survey are peripheral alliances between non-responding microfirms.

Network regressions

Network data violate the assumption of independence between observations and thus require alternative methods to generate test statistics. Hence, regression analysis uses the QAP technique to test for associations with the dependent variable. The quadratic assignment procedure (QAP) calculates the Pearson correlation coefficient for two observed matrices and compares it with a self-generated distribution of correlations based on random permutations of the reference matrix (Kilduff and Krackhardt, 1994; Krackhardt, 1987). All regression analyses were computed with the software package Ucinet (Borgatti et al., 2002).

5. Results

Visualizing network growth in German stock photography

The study of network evolution is promising especially in early stages of network growth, since network structure has not yet become fully institutionalized (Kenis and Knoke, 2002). In the case of stock photo alliances this seems to be the case since it is a more recent reaction to digitization and competitive shift. New agencies and new ties in the alliance network remained quite limited throughout the 1990s, with a clear and disproportional increase since 1999. This increase is reflected in rights-managed (RM) as well as royalty-free (RF) photography and also in the domestic rights-managed alliance network in Germany (Figure 14.1): 46 agencies were founded in the considered period between 1989 and 2005, while 29 had already been established before 1989.

Correlation analysis of firm attribute data supports the argument that digitization and royalty-free photography reduce transactional friction and ease inter-agency cooperation. First, companies with RF-photos receive a higher percentage of pictures from partner agencies and they tend to generate a higher proportion of revenues internationally. At the same time, RF agencies do not differ in terms of firm age or employment size. Second, there is a more general trend that the higher the proportion of digital images, the higher the share of photo supply through sales partners. Technological and institutional change thus fosters the emergence of a new division of labour and a growth in inter-agency sales partnerships.

Figure 14.2 displays four stages of the domestic RM alliance network in five-year intervals between 1990 and 2005. The visualization of this network illustrates the rapid, exponential expansion of alliances since 2000 only. The network started off from a few disconnected cores in 1990 that remained rather stable and small throughout the 1990s. Toward the end of the decade, as the number of new ties increased, one large component emerged in which two thirds of all alliances are connected either directly or indirectly.

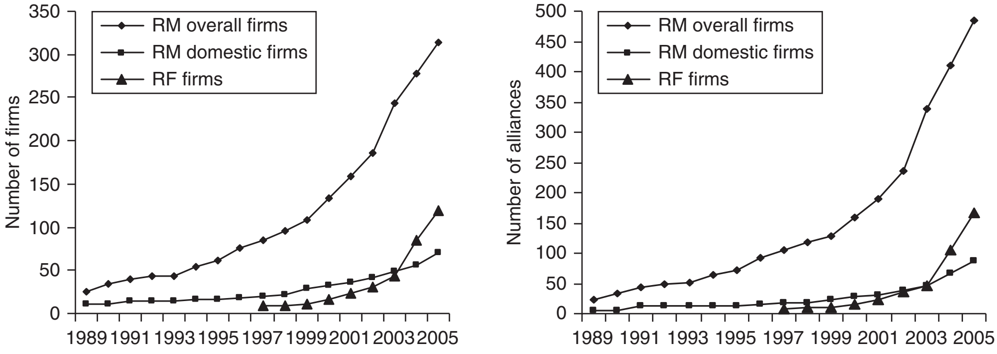


Figure 14.1 Growth of alliance network by members and by licence agreements, 1989–2005

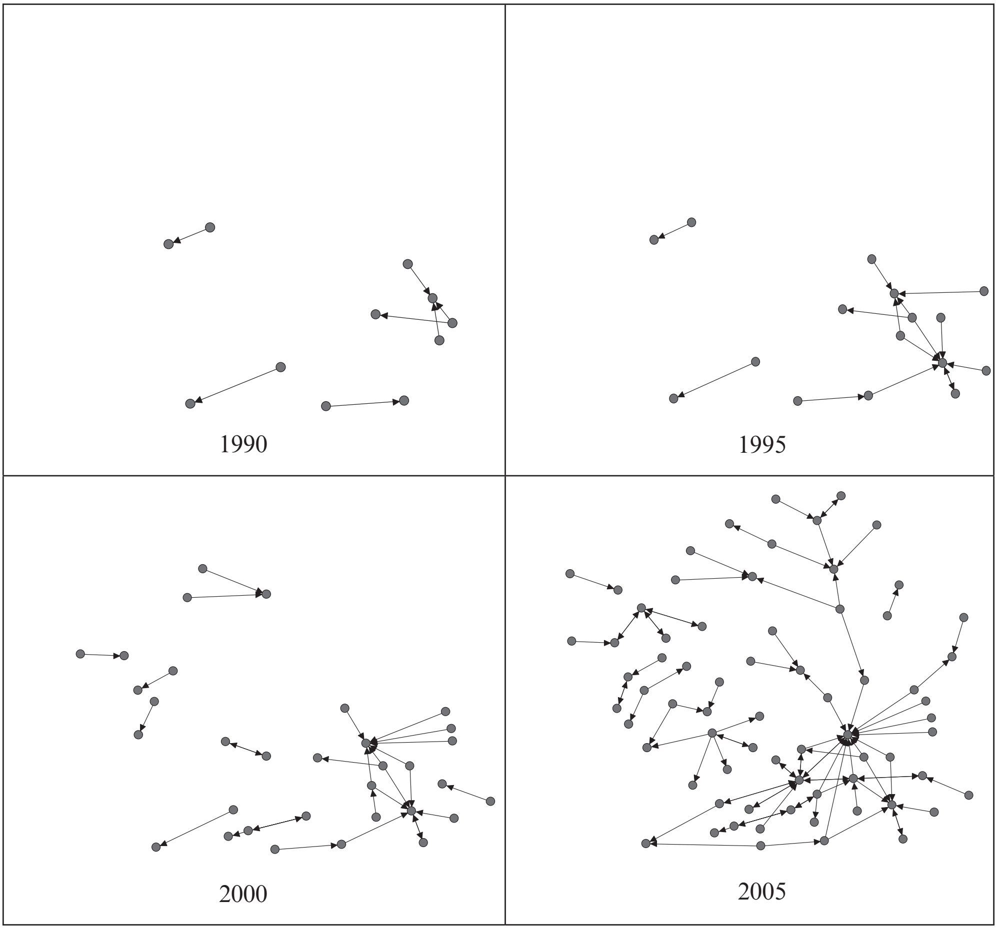


Figure 14.2 Stages of the domestic RM sales alliance network, 1990–2005

Table 14.2 Bivariate QAP regression analysis: all sales alliances in 2005

Independent variables	Regression coefficient	R^2 (adjusted)	Observations
A. Homophily			
Firm age	0.000	0.000	2070
Digital entry	0.002*	0.003	1560
Size employment	0.001*	0.009*	1332
Size images	0.000	0.000	2450
Size photographers	0.001	0.000	1806
Universality	0.002	0.000	2756
Specialization	0.016*	0.002*	2756
International sales	0.000	0.000	1190
Lobby association	0.012**	0.002**	4830
B. Geography			
Colocation	0.023**	0.004**	2652
C. Popularity bias			
Indegree centrality	0.002**	0.006**	4830
Outdegree centrality	0.001*	0.004*	4830
D. Multiconnectivity			
Point connectivity	0.229**	0.220**	4830
Geodesic count	0.207**	0.195**	4830

* $p < 0.05$; ** $p < 0.01$; 2000 random permutations.

One third of alliances however forms a rather peripheral landscape of various unconnected and isolated alliances. In the next analytic section, social network analysis is used to search for an explanation of this emerging network structure.

Explaining tie formation in the rights-managed sales alliance network

This section applies the hypotheses on network growth to the domestic RM alliance network in Germany. Two sets of analyses are carried out. First, in a static perspective the independent variables are tested against the network of sales alliances all in the same year 2005. The results for bivariate QAP regressions are summarized in Table 14.2. This produces a number of significant explanatory variables for further analysis. Second, and in a dynamic perspective, significant variables from the first set of regressions in the year 2005 are used for a year-to-year longitudinal analysis between 1995 and 2005. Each independent variable from one year t_n is compared to the alliance network structure of the subsequent year t_{n+1} thus providing a year-to-year estimation of the hypothesized effect on the new network structure in the subsequent year. The regression effects for the four hypotheses are found to be quite different.

Homophily Arguments about preferential partner selection based on homophily suggest that the more similar two partners are to each other, the more likely will they engage in a relationship. However, this proposition receives hardly any support in the context of the alliance network between picture agencies. Most of the similarity variables are insignificant except for employment size, specialization and business association membership. According to this finding, collecting image material in the same topic specialty and being a member of the business association increases the odds of having a licence agreement. In contrast, however, dissimilarity also drives alliance formation. The higher the difference in employment size, the more likely are two agencies to initiate a licence agreement. This result would argue more in favour of diversity than similarity. It might reflect an emerging division of labour between big agencies that source their image material increasingly from small supplying agencies (see Glückler, 2005). It has to be emphasized that none of the significant effects explains more than 1 per cent of overall variance in the distribution. Hence, the impact of firm similarity may be summarized as being negligible.

Geography A second argument about preferential partner choice holds that geographical proximity between partners increases the likelihood of an alliance. Three quarters of all agencies in the 2005 alliance network are located in one of the top five metropolitan

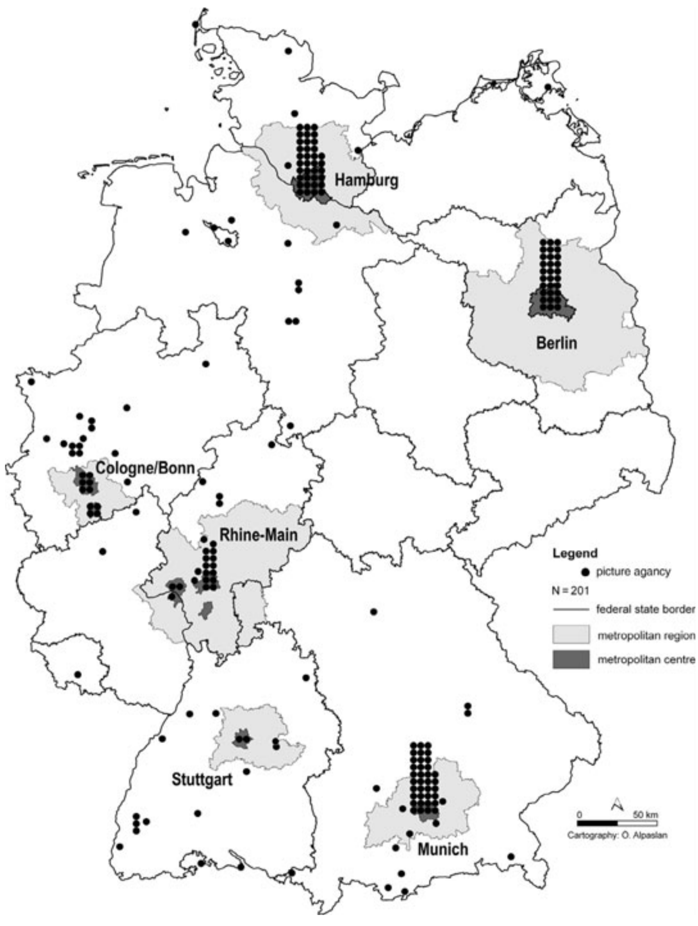


Figure 14.3 Metropolitan clustering of stock picture agencies in Germany, 2005

regions, Hamburg, Munich, Berlin, Frankfurt and Cologne (Figure 14.3). Regression analysis shows that if two firms are located in the same metropolitan region, they are more likely to have a licence agreement. Though the association is highly significant, again, location does not explain more than 0.4 per cent of overall variation and thus can only be an auxiliary condition for alliance formation. The analysis of a series of year-to-year regressions offers an additional and quite remarkable finding. Over the ten-year period since 1995, co-location has become increasingly important in the structuration of alliances. While until 1999, the pattern of domestic sales partnerships was independent from geography, the new ties in the network have been increasingly constrained by geographical propinquity within the dominant metropolitan regions. The positive effect of co-location in the same metropolitan region has grown steadily over the last five years, yet remaining a highly contingent condition for tie formation.

Popularity bias The argument about accumulative advantage predicts that a partner with many previous contacts is more likely to have new ties than a partner with only a few prior relations. Evidence from the regression analysis suggests that the higher the difference in the number of licence agreements between a pair of agencies, the more likely are they to have a relationship. This result applies to both licence-giving and licence-taking, and reflects the overall centralization of alliances on a smaller number of highly connected agencies. In line with the former hypotheses, however, popularity bias accounts for less than 1 per cent of the variance in the distribution of network ties.

Multiconnectivity Finally, the argument of multiconnectivity theorizes that multiple indirect linkages between partners enhance the formation of direct ties between these organizations. In comparison with the other hypotheses, this argument seems best to represent the pattern of licence agreements in the German stock photography market. Point connectivity as well as geodesic count are highly significant and explain around 20 per cent of the variance in the distribution of sales alliances. Both measures particularly focus on unconnected firms in the network and develop notions of the quality of their indirect interconnection through third parties. If there are many different intermediate contacts between any two firms, and if there are many alternative shortest connections between the two, they can be considered as being multiply connected (multiconnectivity) – although there is no direct link between them. The analysis of a series of year-to-year regressions reinforces this association. The point connectivity and geodesic count for each pair of agencies in year t_n explain roughly between 40 per cent and 60 per cent of the network structure in each subsequent year t_{n+1}. In addition, the ten-year time series since 1995 conveys that although both measures are continuously significant, their explanatory value has suffered in the last couple of years because more and more new ties enter the alliance network and challenge the existing structure (Table 14.3). One conclusion might be that when networks face major expansion and high rates of new tie formation, the incumbent network structure imposes less constraint on new ties and may experience a period of path-destruction.

6. Discussion and conclusion

This chapter has sought to combine research on network evolution with an evolutionary approach to economic geography. A number of different theoretical arguments about

Table 14.3 R^2s for year-to-year QAP regressions of the alliance network, 1995–2005

	New ties	All ties	Location†	Point connectivity	Geodesic count
1995 to 1996	2	16	0.000	0.519**	0.638**
1996 to 1997	1	17	0.000	0.536**	0.654**
1997 to 1998	1	18	0.000	0.552**	0.668**
1998 to 1999	4	22	0.000	0.489**	0.587**
1999 to 2000	7	29	0.002*	0.461**	0.519**
2000 to 2001	3	32	0.003*	0.609**	0.655**
2001 to 2002	7	39	0.003*	0.544**	0.608**
2002 to 2003	8	47	0.004**	0.526**	0.607**
2003 to 2004	19	66	0.002*	0.428**	0.459**
2004 to 2005	21	87	0.004**	0.158**	0.123**

Notes:

* $p < 0.05$, ** $p < 0.01$, 2000 random permutations; R^2s are adjusted. Between 1989 and 1995 there were only eight new entries and a total of 14 alliances. Year-to-year regression would not convey interpretable results before 1995.

† The information on the location of an agency is only available for 2005 and thus used consecutively for regressions in each year of alliance network.

the nature of network expansion were discussed and empirically applied to the context of inter-organizational alliances in stock photography. Although stock photography is a small market in financial terms and has been largely ignored in academic research, it is a particularly interesting organizational field for the study of network dynamics. The market has experienced fundamental changes with respect to various dimensions: digitization in technological terms, new licence models in institutional terms, and new organizational practices as well as major new entrants to the incumbent field. It has been argued that these environmental changes have given rise to new forms of sales alliances between picture agencies. Despite the fact that agencies renounce a significant share of revenue with a distribution partner, they increasingly now engage in domestic partnerships. Moreover, the sheer number and rate of new alliance formation has grown exponentially over the last couple of years. Since this phenomenon is relatively recent and still in an early phase, the evolving alliance network represents an ideal case for the study of network evolution.

The analysis revealed that organizational homophily and popularity bias are relatively weak conditions for estimating the locus of new alliances. Only organizational diversity, that is, difference in firm size, and geographical co-location were highly significant, yet relatively weak predictors for network growth. The multiconnectivity hypothesis (Powell et al., 2005) received stronger support. It was measured as the extent to which two agencies were multiply and indirectly connected via alternative third parties. Multiconnectivity explained on average half of the variance in the distribution of alliances across the network over a consecutive ten-year period. In sum, two agencies were more likely to form a sales partnership if they were located in the same metropolitan region, had different firm size, and if they were indirectly connected through other organizations in the field. On the other hand, while over time the tendency to cooperate locally has increased, the impact of multiple linkages on the alliance network has decreased. This is mainly because in recent years, many new agencies were established and formed many new alliances in the network. Given the limit of this study on stylized survey data it is difficult to interpret the emerging effect of geographical propinquity on new tie formation appropriately. Although the picture trading business has become mostly virtual, the new uncertainties emerging from the dynamic market environment seem to facilitate strategic partnerships in spatial proximity. The local milieu or creative fields (Scott, 2000) might be argued to represent both an imprinted institutional mode of communication among the creative professionals in a city or city region (Marquis, 2003), and a territorialized form of culture (Maillat et al., 1995) that conveys positive communication externalities and reduces uncertainty through more monitorable interaction. Though the association between geography and alliance behaviour corresponds well with other empirical evidence (Powell et al., 2005), more detailed research on the nature of sales alliances is necessary to support the findings in the context of stock photography.

It is an open question for further research whether and under what conditions rapid network expansion reproduces an imprinted structure over long times to lock-in as a path of network evolution or whether it disrupts established patterns and creates a new path to network growth (Martin and Sunley, 2006). The study presented here is only a first and exploratory analysis toward a currently emerging field of evolutionary theory on networks. For economic geography, this is a promising line of reasoning not only because geography learns from their input but also because it may essentially contribute to the development of solutions. If geography is a continuous constraint (or enabling condition) for networks to form, how then do networks travel over space in the course of their development and how do regions anchor these networks to retain their dynamics locally?

The focus of this chapter has been on the mutual effects that network structure imposes on future tie formation and, vice versa, the effect of new ties on the existing network structure. For this purpose, the research strategy has focused on a more standardized, quantitative approach to these patterns of variation over time. The specific regional context that reflects in a set of interrelated institutions, the particular set of actors and a given resource endowment will also play an important role for the understanding of network formation in a geographical perspective. Future research should dedicate more enquiry also to institutional aspects of geographical network evolution.

References

Albert, P. and G. Feyel (1998), 'Fotografie und Medien. Die Veränderungen der illustrierten Presse' [Photography and media. The change in the illustrated press], in M. Frizot (ed.), *Neue Geschichte der Fotografie* [New History of Photography], Köln: Könemann, pp. 358–69.

Baum, J.A., A.V. Shipilov and T.J. Rowley (2003), 'Where do small worlds come from?', *Industrial and Corporate Change*, **12**, 697–725.

Borgatti, S.P., M.G. Everett and L. Freeman (2002), *Ucinet 6 for Windows*, Harvard, CT: Analytic Technologies.

Boschma, R.A. and K. Frenken (2006), 'Why is economic geography not an evolutionary science? Towards an evolutionary economic geography', *Journal of Economic Geography*, **6**, 273–302.

Boschma, R.A. and S. Iammarino (2009), 'Related variety, trade linkages and regional growth in Italy', *Economic Geography*, **85** (3), 289–311.

Boschma, R.A. and J.G. Lambooy (1999), 'Evolutionary economics and economic geography', *Journal of Evolutionary Economics*, **9**, 411–29.

Brass, D.J., J. Galaskiewicz, H.R. Greve and W. Tsai (2004), 'Taking stock of networks and organizations: a multilevel perspective', *Academy of Management Journal*, **47**, 795–817.

CEPIC (2001), *CEPIC European Industry Survey 2001*, Berlin: CEPIC.

Fleming, L. and O. Sorenson (2001), 'Technology as a complex adaptive system: evidence from patent data', *Research Policy*, **30**, 1019–39.
Frizot, M. (1998), '1838–1840. Fotografische Entdeckungen' [Photographic discoveries], in M. Frizot (ed.), *Neue Geschichte der Fotografie* [New History of Photography], Köln: Könemann, pp. 22–31.
Frosh, P. (2001), 'Inside the image factory: stock photography and cultural production', *Media, Culture and Society*, **23**, 625–46.
Glückler, J. (2005), 'Digitalisierung und das Paradox informatorischer Reichweite in der Agenturfotografie' [Digitisation and the information range paradox in the photography agency], *Geographische Zeitschrift*, **93**, 100–20.
Glückler, J. (2007), 'Economic geography and the evolution of networks', *Journal of Economic Geography*, **7**, 619–34.
Goddard, L. (1995), 'Focus photo libraries: pictures at the touch of a button', in *PR Week*, London.
Hite, J.M. and W.S. Hesterly (2001), 'The evolution of firm networks: from emergence to early growth of the firm', *Strategic Management Journal*, **22**, 275–86.
Kenis, P. and D. Knoke (2002), 'How organizational field networks shape interorganizational tie-formation rates', *Academy of Management Review*, **27**, 275–93.
Kilduff, M. and D. Krackhardt (1994), 'Bringing the individual back in: a structural analysis of the internal market for reputation in organizations', *Academy of Management Journal*, **37**, 87–108.
Kilduff, M. and W. Tsai (2003), *Social Networks and Organizations*, London: Sage.
Kjemtrup, I. (1997), 'ASPP photo editor/researcher survey. Part 1', *The Picture Professional (American Society of Picture Professionals)*, available at: www.aspp.com.
Knoke, D. (2001), 'Making connections', in D. Knoke (ed.), *Changing Organizations: Business Networks in the New Political Economy*, Boulder, CO: Westview Press, pp. 120–63.
Krackhardt, D. (1987), 'QAP partialling as a test of spuriousness', *Social Networks*, **9**, 171–86.
Lavie, D. (2004), 'The evolution and strategy of interconnected firms: a study of the UNISYS alliance network', *Academy of Management Proceedings*, E1–E6.
McPherson, M., L. Smith-Lovin and J. Cook (2001), 'Birds of a feather: homophily in social networks', *Annual Review of Sociology*, **27**, 415–44.
Maillat, D., B. Lecoq, F. Nemeti and M. Pfister (1995), 'Technology district and innovation: the case of the Swiss Jura', *Regional Studies*, **29**, 251–63.
Marquis, C. (2003), 'The pressure of the past: network imprinting in intercorporate communities', *Administrative Science Quarterly*, **48**, 655–89.
Martin, R. and P. Sunley (2006), 'Path dependence and regional economic evolution', *Journal of Economic Geography*, **6**, 395–437.
Owen-Smith, J. and W.W. Powell (2004), 'Knowledge networks as channels and conduits: the effects of spillovers in the Boston biotechnology community', *Organization Science*, **15**, 5–21.
Powell, W.W., K.W. Koput and L. Smith-Doerr (1996), 'Interorganizational collaboration and the locus of innovation: networks of learning in biotechnology', *Administrative Science Quarterly*, **41**, 116–45.
Powell, W.W., D. White, K.W. Koput and J. Owen-Smith (2005), 'Network dynamics and field evolution: the growth of interorganizational collaboration in the life sciences', *American Journal of Sociology*, **110**, 1132–205.
Rich, M. (1994), 'The Queen's Awards 1994: the picture of success – Tony Stone Images sells photographs around the world', *The Financial Times*, London.
Rigby, D.L. and J. Essletzbichler (2006), 'Technological variety, technological change and a geography of production techniques', *Journal of Economic Geography*, **6**, 45–70.
Rowley, T., D. Behrens and D. Krackhardt (2000), 'Redundant governance structures: an analysis of relational and structural embeddedness in the steel and semiconductor industries', *Strategic Management Journal*, **21**, 369–86.
Rowley, T.J., H.R. Greve, H. Rao, J.A.C. Baum and A.V. Shipilov (2005), 'Time to break up: social and instrumental antecedents of firm exits from exchange cliques', *Academy of Management Journal*, **48**, 499–520.
Sachs, C.D.P. (2003), 'ASPP 2003 members survey', *The Picture Professional*, 31–2.
Scott, A.J. (2000), *The Cultural Economy of Cities*, London: Sage.
Sorenson, O. (2003), 'Social networks and industrial geography', *Journal of Evolutionary Economics*, **13**, 513–27.
Sorenson, O. (2005), 'Social networks, informational complexity and industrial geography', in D.B. Audretsch, D. Fornahl and C. Zellner (eds), *The Role of Labour Mobility and Informal Networks for Knowledge Transfer*, New York: Springer, pp. 79–95.
Staber, U. (1997), 'An ecological perspective on entrepreneurship in industrial districts', *Entrepreneurship and Regional Development*, **9**, 45–64.
Staber, U. (2001), 'The structure of networks in industrial districts', *International Journal of Urban and Regional Research*, **25**, 537–52.

Stinchcombe, A.L. (1965), 'Social structure and organization', in J.G. March (ed.), *Handbook of Organizations*, Chicago: Rand McNally, pp. 142–93.
Storper, M. and A.J. Venables (2004), 'Buzz: face-to-face contact and the urban economy', *Journal of Economic Geography*, **4**, 351–70.
Walker, G., B. Kogut and W. Shan (1997), 'Social capital, structural holes and the formation of an industry network', *Organization Science*, **8**, 109–25.
Wilkinson, H. (1997), '"The new heraldry": stock photography, visual literacy, and advertising in 1930s Britain', *Journal of Design History*, **10**, 23–38.
Zucker, L.G. (1986), 'Production of trust: institutional sources of economic structure, 1840–1920', *Research in Organizational Behavior*, **8**, 53–111.

15 Complexity, networks and knowledge flow*

Olav Sorenson, Jan W. Rivkin and Lee Fleming

1. Introduction

Though not framed as such, this chapter emerges directly from an evolutionary perspective on economic geography. To provide some sense of its relationship to the rest of the volume, we begin with a brief account of the path leading up to this study. After providing this context, we then discuss the chapter's more immediate implications for the literature on evolutionary economic geography.

The journey began with a study of entrepreneurs in the multimedia industry. Though never published, interviews with and surveys of firm founders led the first author to question the assumption that entrepreneurs actively evaluate and choose locations. Instead, they seem more likely simply to found their firms in the regions in which they reside. In part, this behavior undoubtedly stems from a desire to avoid the costs of moving. But it also reflects the difficulty of starting a new venture, particularly outside of one's home region. Social connections help entrepreneurs to identify opportunities, to raise financial capital and to recruit employees. For most, the relations that enable these processes remain firmly rooted in their local regions, and therefore they tend to bind entrepreneurs to the regions in which they reside.

Building on this assumption and the fact that a large proportion of entrepreneurs begin businesses in industries in which they have prior experience, the first author proposed that industries should agglomerate even when collocation offered no benefits. Since those with experience in an industry tend to reside in close proximity to existing firms in the industry (where they presumably gained their experience), entrepreneurs in any industry should emerge at higher rates in regions with dense concentrations of existing enterprises in the industry, leading to increasing agglomeration. With several coauthors, the first author has demonstrated that this process appears to account for the geographic concentration of footwear manufacturers (Sorenson and Audia, 2000), biotechnology firms (Stuart and Sorenson, 2003) and computer workstation manufacturers (Sorenson, 2005) in the United States, and on average across all types of manufacturers and service providers in the Danish economy (Dahl and Sorenson, 2009).

This general theory of the geographic concentration of industries fits very much within the domain of evolutionary economic geography. It satisfies the key criteria proposed by Boschma and Martin (2007) for such a perspective. It provides an explanation for the changing patterns of economic activity over time, and this explanation implies path dependence, since the choices of past entrepreneurs inexorably affect the availability of potential future entrepreneurs in each region.

This chapter connects to this larger theory by addressing the question of when these processes should reinforce the existing geographic distribution of an industry most strongly. Not all industries agglomerate to the same degree. Some concentrate in a few regions; others disperse more broadly. Though other factors certainly influence these differing patterns of development, with respect to the theory outlined above, a key issue in determining the degree to which an industry disperses is the extent to which existing social connections within the industry prove advantageous to entrepreneurs. In this chapter, originally published in *Research Policy*, we argue that the complexity of the underlying knowledge importantly influences when a connection to the source of the knowledge proves most valuable. In particular, relationships offer the greatest advantage in accessing knowledge of intermediate interdependence.

The logical implication of this idea for the geographic evolution of industries is that those industries that rely on technology and processes of intermediate complexity should remain most concentrated in space. Those based on less complex knowledge, meanwhile, should diffuse more readily to new locations. Though from the theory one might also expect greater geographic dispersion in industries dependent on highly complex technology and processes, it seems quite likely that such industries fail to emerge, as the transfer of the knowledge proves so difficult as to preclude its replication across firms. Though this proposition has yet to receive rigorous testing, Sorenson (2004) shows that it is consistent with the cross-sectional variation in the geographic concentration of industries in the United States in the 1990s.

Though clearly of relevance to an evolutionary perspective on economic geography, the processes described in the chapter also have implications beyond geography. Given the range of potential applications, we chose to frame the original study more broadly.

2. Complexity, networks and knowledge flow

Original introduction

The flow of knowledge plays a central role in a wide variety of fields (for a review, see Rogers, 1995). Sociologists began investigating diffusion processes – and the importance of social structure to those processes – to understand the adoption patterns of agricultural and medical innovations (Coleman et al., 1957; Ryan and Gross, 1943). To students of technology management, knowledge flow first arises as an important issue in the context of technology transfers within the firm (Allen, 1977; Teece, 1977), but questions of diffusion also arise when technology scholars ask whether incumbent firms or upstarts first develop and commercialize new inventions (Reinganum, 1981; Tushman and Anderson, 1986). Both students of organizational learning (for a review, see Argote, 1999) and industrial economists (Griliches, 1957; Irwin and Klenow, 1994; Zimmerman, 1982) study how knowledge moves through firms and how it spills over to other firms. In short, a diverse array of scholars shares an interest in knowledge diffusion processes.

The normative interpretation given to diffusion, however, differs dramatically across fields. Economists and sociologists tend to focus on the societal benefits of spillovers (i.e. the flow of knowledge across actors, usually firms). The generation of new knowledge often requires substantial investment in research and development, but the repeated application of this knowledge, once produced, entails little if any incremental cost (Arrow, 1962). Knowledge diffusion, therefore, engenders scale economies and stimulates economic development by allowing several firms to benefit from the R&D activities undertaken by a single firm (Marshall, 1890; Romer, 1987; Scherer, 1984). Management scholars, by contrast, note that when knowledge escapes to competing firms the returns to innovation become fleeting at best. As rivals imitate new products and processes, the degree of differentiation or cost advantage accruing to the innovator erodes. The business literature thus urges managers to defend against spillovers (Kogut and Zander, 1992; Lippman and Rumelt, 1982).

Though their prescriptions differ, economists, sociologists, strategists, and students of technology management all seek a better understanding of why some knowledge disperses widely while other knowledge does not. In this quest, some scholars have focused on the attributes of the knowledge itself. For example, highly specific knowledge may flow slowly because few parties other than the initial innovator either have the baseline knowledge and skills necessary to absorb it (Cohen and Levinthal, 1990) or can benefit from its application (Henderson and Cockburn, 1996; McEvily and Chakravarthy, 2002). Other studies focus on how social networks structure the flow of knowledge (e.g. Coleman et al., 1957; Hansen, 1999; Singh, 2005), implicitly attributing the rate of diffusion to the locus of innovation in the network.

This chapter seeks to augment our understanding of knowledge flow by examining the interplay between two features: social proximity and the complexity of the underlying knowledge.[1] Social proximity here refers to the distance between two parties in a social network; for example, one would consider those who have a direct relationship to each other to be closer than those who have a mutual acquaintance but have never met. We meanwhile define complexity in terms of the level of interdependence inherent in the subcomponents of a piece of knowledge (Kauffman, 1993; Simon, 1962; cf. Zander and Kogut, 1995). Interdependence arises when a subcomponent significantly affects the contribution of one or more other subcomponents to the functionality of a piece of knowledge. When subcomponents are interdependent, a change in one may require the adjustment, inclusion or replacement of others for a piece of knowledge to remain effective.

Consider then an actor who is a source of knowledge and two potential recipients of that knowledge – one socially close to the source and one further away. When does the proximate actor have the greatest advantage over the distant in receiving and building on the knowledge? We argue that *the advantage should peak when the underlying knowledge is of moderate complexity*. Our expectation emerges from the recognition that receiving and building on knowledge frequently requires the recipient to engage in search to fill in gaps and correct transmission errors in the knowledge conveyed – the cost and difficulty of which increase with knowledge complexity. Social proximity reduces the need for search by facilitating high-fidelity transmission (i.e. complete information with negligible noise). On the other hand, as the social distance separating the source and the would-be receiver grows, unaided search plays an increasingly important role in diffusion. Under such conditions, simple knowledge should flow universally – to actors near and far – because search can easily substitute for high-fidelity transmission. Highly interdependent knowledge meanwhile defies diffusion, regardless of whether one relies on search or social proximity. For knowledge of moderate complexity, however, a gap emerges between the ability of close actors, relative to that of distant actors, to receive and build on knowledge. High-fidelity transmission gives proximate actors sufficient insight that they can succeed in receiving and building on knowledge, even where more distant actors, who rely more heavily on search, fail.

We analyze patent data to test our thesis empirically. Citation patterns across patents offer something of a fossil record for the flow of knowledge – providing a lasting reflection of ephemeral interactions. Using this record, we estimate the effect of knowledge complexity on the likelihood of future citations as a function of the social proximity of future inventors to the inventor of the original piece of knowledge, comparing those socially close to and far from the source. To assess social proximity, we calculate the geodesic length between patents' inventors in a collaboration network. We also supplement this metric with indicators of geographic proximity and employment within the same organization. To gauge complexity, we develop a measure that reflects the historical interdependence of a patent's subcomponents with other subcomponents. The findings provide strong support for our core hypothesis: the higher likelihood of citation among proximate inventors peaks for knowledge of an intermediate level of complexity (interdependence).

This work contributes to the literature in several ways. First, from the perspective of social networks, it identifies one condition under which social proximity should prove especially important to knowledge flow: for knowledge of intermediate complexity. Though social scientists have usefully demonstrated that networks matter for the diffusion of knowledge, relatively little research considers precisely when those networks should matter most (Baker and Faulkner, 2004; Strang and Soule, 1998). By synthesizing the social network perspective with work on conceptions of knowledge receipt as search, we identify scope conditions on the relevance of social connections to the diffusion process. Second, with respect to evolutionary economics, our work highlights social connections as an important channel through which 'insiders' gain superior access to knowledge. Extant work asserts that insiders – defined usually as those within the same firm as the source – have better access to an original success, which serves as a template in efforts to transfer and extend that knowledge (Nelson and Winter, 1982, p. 119; Rivkin, 2001). Yet this work fails to establish the source of this preferential access. Does it come from incentives that reward transfer, from the confidentiality agreements that employees sign, or from some other source? Our research points to direct social connections as a critical factor differentiating these internal parties from those outside the firm.

3. The flow of complex knowledge

Our discussion begins with the most common finding of classic diffusion studies: the S-shaped cumulative adoption curve (Griliches, 1957; Ryan and Gross, 1943; and Rogers, 1995, provides an excellent review). Researchers consistently find that the adoption of an innovation over time follows a common pattern: growing slowly at first, then accelerating rapidly, and finally slowing to reach some asymptotic saturation level. These dynamics resemble that of an epidemic spreading through a population; the innovation first 'infects' those most at risk of exposure – actors closest to the original source (Hägerstrand, 1953 [1967]) – and those most susceptible to infection – those most prepared to accept the uncertainty associated with an untested technology (Mansfield, 1968) or whose idiosyncratic characteristics make the innovation appear most attractive (Griliches, 1957). Over time, awareness of the innovation spreads, uncertainty ebbs, and the economics of the invention become favorable to a larger share of the population. Diffusion then takes off. In this classic perspective, new knowledge resembles a stone thrown into a calm pond, its ripples moving steadily across the entire surface.

Though this pattern accurately describes the diffusion of a wide variety of innovations and knowledge, critics have faulted this focus on the S-curve for several reasons (see Mahajan et al., 1990; Hargadon, 1998). Two of these critiques have particular relevance here. First, the classic diffusion literature typically depicts knowledge as moving unaltered as it passes from one actor to the next. Contrary to this depiction, in reality transmission rarely occurs with perfect fidelity. Both gaps in the information sent and errors in its interpretation typically require the receiver to reconstruct portions of the original knowledge. This process occurs so commonly that it even forms the basis of amusement in the children's game of telephone.[2] Most knowledge, therefore, requires effort to acquire and transmutes to some extent as actors strive to receive and build on it; recipients assimilating new knowledge must actively process it by experimenting with its application to new problem domains and environmental contexts. Witness, for instance, the efforts of American automakers as they struggled to digest the knowledge embodied in Japanese lean production techniques (Womack et al., 1990) or the labors of computer makers as they sought to imitate Dell's direct distribution model (Porter and Rivkin, 1999). In both cases, the receipt of knowledge required years of trial, error, reflection, and adjustment and, arguably, remains incomplete.

Even within the supportive infrastructure of an organization, receiving and building on new knowledge can prove difficult. Teece (1977), for example, reports that the transmission and assimilation of technical know-how accounted for 19 percent of project costs, on average – running as high as 59 percent in one case – in 26 international technology transfer projects. Chew et al. (1990) find the internal transfer of best practices so incomplete in multi-plant commercial food operations that, within a firm, the best plants produce twice as efficiently as the worst, even after controlling for differences in processing technology, location, and plant size (Szulanski, 1996, offers additional evidence). Hence, we regard the act of receiving and building on knowledge not as the acceptance of a complete, well-packaged gift, but rather as the beginning of a trial-and-error process.

Our second concern regarding the simple S-curve characterization of diffusion arises from its inattention to the crucial role that social networks play in diffusion. Several studies, largely out of sociology, demonstrate that knowledge spreads from its source not in concentric circles, but along conduits defined by social connections (Burt, 1987; Coleman et al., 1966; Lazarsfeld et al., 1944; see Marsden and Friedkin, 1993, for a review). Consider some of the relevant findings: Hedström (1994) discovered that network density and geographic proximity can explain most of the spread of the idea of unionization in Sweden. In an analysis of adoption patterns for 'poison pills' and 'golden parachutes', Davis and Greve (1997) offered strong evidence that information about these policies traveled through corporate board interlocks. And Hansen (1999) found that strong ties best conveyed complex knowledge across product development teams within a firm. A growing literature thus points to the importance of social networks as pathways that channel the flow of knowledge among actors.

We synthesize these two perspectives – knowledge receipt as an active process of experimentation and search, and an appreciation for the role of social networks – into a model of knowledge flow. The model offers unique predictions regarding how knowledge complexity influences patterns of success among efforts to receive and extend knowledge.

Knowledge receipt as search

Building on the intellectual scaffolding of evolutionary economics, our perspective conceptualizes a piece of knowledge as a recipe (Nelson and Winter, 1982).[3] The list of potential ingredients encompasses both physical components and processes. The recipe details how to combine these ingredients – in which proportions, in what order, in what circumstances – to achieve a desired end. For instance, a recipe for a McDonald's outlet might read something like: 'When a customer places a special order, the counter clerk keys the order into the register, which causes the order to show up on the computer screen in the kitchen, which induces the cook to put a raw hamburger on the grill. . .' Or: 'when opening a new outlet, a manager in the real estate department secures a site while the franchising office identifies a franchisee. Next, the franchisee contacts construction contractors while hiring shift managers. . .' Though these recipes may appear in writing, they more commonly reside in the form of behavioral routines, individual memory, or technology (March and Simon, 1958).

The conceptualization of knowledge as a recipe leads naturally to thinking of innovation as a process of searching for new recipes. Following a long tradition (Gilfillan, 1935; Schumpeter, 1939; Usher, 1954), Nelson and Winter (1982) explicitly treat innovation as a search process; inventors explore the space of possible combinations of ingredients, or recipes, for new and better alternatives. This exploration involves not just the search for the best combinations of ingredients but also the quest for the most effective methods of integrating them. Researchers who conceptualize innovation as search frequently exploit a landscape metaphor as a means of providing an intuitive understanding of the search process (Fleming and Sorenson, 2001; Levinthal, 1997; Rivkin, 2000). Innovators – depicted as myopic in their awareness of the terrain – search these landscapes for peaks, which represent good recipes or useful inventions.

Once a useful innovation has been located, transferring its recipe, even between cooperative actors, can fail for two reasons. First, the recipient rarely grasps the original recipe completely, due to imperfections in the transfer process. Gaps emerge in what the sender conveys – perhaps the chef forgets an ingredient or skips a step – and the receiver may misinterpret some of the information that is transmitted. And, unless the recipient understands *perfectly* the recipe that generated the success – an unlikely situation – she must engage in search to fill the gaps and correct the errors in her version of the recipe. Any attempt to receive and extend a recipe in new settings will likewise require the recipient to rediscover the original combination, or some variant of it better suited to the new context.

Second, the local ingredients and cooking experience of the receiving chef rarely match identically those of the sender. Research on absorptive capacity (Cohen and Levinthal, 1990) emphasizes that successful knowledge diffusion requires the receiver to possess a base of knowledge and skills to assimilate new information. Without this baseline, the transmission of new discoveries would often entail the communication of exorbitant amounts of data; imagine how long a recipe would become if one needed to detail every step of the process – how to chop vegetables, how to boil water, and so on. These two factors imply that knowledge recipients rarely, if ever, act merely as passive beneficiaries; they actively search, recreate, and build upon the original recipes.

In this process, certain types of recipe prove particularly tricky to transfer because the sender finds it difficult to specify and communicate precisely where the original combination resides in the combinatorial space of ingredients; on the figurative treasure map, it is hard to place the 'X' that 'marks the spot'. This communication difficulty could arise as a result of causal ambiguity (Lippman and Rumelt, 1982; Reed and DeFillippi, 1990): the innovator might not fully understand the connection between actions and outcomes so the roots of the original success remain unclear. It could also occur because the production process calls on tacit personal skills or connections among individuals that the involved parties themselves do not consciously understand (Polanyi, 1966; von Hippel, 1988), or that eludes codification (Zander and Kogut, 1995). These factors essentially increase the likelihood that the knowledge transmitted has gaps. The complexity of the recipe itself can also impair knowledge flow by increasing the difficulty for the recipient of filling these gaps and correcting transmission errors.

As noted above, complexity refers to the degree to which the components in a recipe interact sensitively in producing the desired outcome. Our definition here closely follows Simon (1962), who classifies a piece of knowledge as complex if it comprises many elements that interact richly (see also Kauffman, 1993; cf. Zander and Kogut, 1995). We adopt Simon's definition, but pay particular attention to the intensity of interdependence among the ingredients in the recipe. A high degree of interdependence indicates that many ingredients influence the effectiveness of others so that a change in one may dramatically reduce the usefulness of the recipe. Replicating the functionality of the original recipe often requires adjustments in the set of other ingredients or the processes for combining them. Low interdependence implies small cross-component effects and a corresponding opportunity to adapt and change ingredients independently.

Discovering, or rediscovering, a complex piece of knowledge poses a stiff challenge. Interdependence produces two effects that undermine the recipient's attempts to receive and build on the original. First, small errors in reproduction cause large problems when ingredients cross-couple in a rich manner. In highly interdependent systems, implementers often realize no value from adopting a set of practices unless each and every component fits into place perfectly; a single error threatens the effectiveness of the entire system. An American automaker that attempts to adopt lean production techniques, for instance, may alter its human resource practices and inventory policies, yet see no benefit because it failed to invest appropriately in flexible production equipment. The fragility of such tightly coupled systems has been well documented (Perrow, 1984; Weick, 1976). Second, interdependence leads to a proliferation of 'local peaks'. These internally consistent – though not necessarily optimal – ways of combining ingredients elude improvement through incremental search because altering any single element degrades the quality of the outcome (Kauffman, 1993). Such local peaks would pose no problem to omniscient actors, who could assess the entire space of possibilities, but for individuals with finite cognitive abilities and a limited purview of the landscape, such search proves difficult; in the face of high interdependence, searchers frequently find themselves trapped on local peaks. Moreover, these local peaks tend to correspond to poor recipes precisely when interdependence creates a thick web of potentially conflicting constraints.

Complexity and access to a template

Success in receiving and building on complex knowledge depends crucially on access to the original success, which serves as a *template* (Nelson and Winter, 1982, pp. 119–20; Winter, 1995). For reasons explored below, individuals differ in their access to the template. Superior access facilitates the knowledge recipient's search in at least two ways. First, the recipient begins searching in closer proximity to the ultimate target – as a result of either fewer errors in the interpretation of the transmission or smaller gaps in the information sent. Second, superior access allows the recipient to solicit advice when problems arise, helping the recipient to home in on the desired knowledge more efficiently.

Consider two actors both trying to receive and build on a valuable piece of knowledge but who differ in their access to the template. The first has superior, though admittedly still imperfect, access to and understanding of the original, successful recipe. The second has far poorer access. To what degree does the first actor's superior but imperfect access to the template have value, in the sense that it enables the actor to receive and build on the original recipe more effectively? We contend that the value of this access depends on the complexity of the underlying knowledge in an inverted U-shaped relationship; that is, intermediate levels of interdependence maximize the value of preferential access.

Suppose first that the ingredients of the knowledge do not interact; getting one element in the recipe wrong diminishes that component's contribution to the whole, but it does not undermine the other components. In this situation, the first actor's access to the template does not educe a persistent advantage. Through routine, incremental search efforts, the second actor can reconstruct the recipe. Few local peaks threaten to trap the poorly informed recipient. As a result, both actors eventually fare equally well; search on the part of a recipient can easily substitute for high-fidelity transmission.

Next consider knowledge with an intermediate degree of interdependence. Local peaks now appear, but they remain relatively few in number. The well-informed actor begins its search near, but not precisely at, the original combination of ingredients. Through incremental search, and with recourse to the template, it can assemble the proper combination of ingredients. The second actor, who likely begins search farther from the target and receives less guidance about the direction in which to explore, more likely becomes ensnared on some local peak, away from and inferior to the original success. Here superior access to the template gives the first actor an advantage that the second cannot recreate through search.

Finally, imagine a piece of maximally interdependent knowledge: ingredients depend on one another in an extremely delicate way, and none produces much benefit unless all align perfectly. Local peaks now pervade the landscape and neither actor's incremental search will likely reproduce or build on the original knowledge with any success. The first actor's superior access to the template thus has little value beyond the second's highly imperfect access.

Taken together, these arguments imply that the advantage of superior but imperfect access to the template reaches its peak at moderate levels of interdependence between knowledge components. With moderate interdependence, the smoothness of the landscape allows a party that begins its search near the desired peak to rediscover it through local search. Yet the landscape also has sufficient ruggedness that a party that begins search far from the target likely finds itself trapped on a lower peak. In contrast, the single-peaked landscape that comes with independent components allows both parties to succeed in receiving and building on the source knowledge through local search. The highly rugged landscape produced by extreme interdependence meanwhile stymies both parties thoroughly. (For a more formal treatment, see the simulation in the Appendix.)

Social networks and template access

The quality of an actor's access to the template may depend on many factors. One crucial factor is the nature of the actor's social relations, which provide conduits through which valuable information travels (Hägerstrand, 1953; Homans, 1950). In particular, we claim that the quality of an actor's access to a template declines with social distance – that is, the number of nodes that separate the actor from the source of the knowledge in a social network. Direct, single-step connections provide the most obvious and valuable links between inventors and those attempting to receive and build on knowledge because they permit two-way communication. The recipient can therefore interactively query the original source of the knowledge to correct errors or to fill gaps in the original transmission.[4]

Short, indirect paths – for example with one or two intervening steps – can also provide beneficial access to the template, as even second-hand information provides important clues about how to reconstruct and build on new knowledge. Mutual acquaintances may also allow for direct communication with the source if they will introduce and vouch for a potential knowledge recipient (Burt, 1992). Moreover, actors removed by only a few steps from the knowledge source will share more background knowledge, a larger proportion of specialized language, and a wider range of beliefs with the source (for a review, see McPherson et al., 2001). All of these facilitate high-fidelity transmission (Arrow, 1974; Cohen and Levinthal, 1990; Durkheim, 1912). The quality of template access, however, undoubtedly declines rapidly as the number of actors between the innovator and the would-be receiver increases; as in the game of telephone, each step in the path between the two parties offers an opportunity for errors and omissions to creep into the transmission.

The previous subsection argued that superior access to the template creates the greatest advantage in knowledge diffusion with knowledge of intermediate complexity (interdependence). Combining that idea with the notion that social proximity provides superior access to the template, we arrive at the central proposition of our chapter:

Hypothesis: In attempts to receive and build on knowledge, actors who are socially close to the source of the knowledge have the greatest advantage over distant actors when the knowledge is of intermediate interdependence.

In sum, we view knowledge diffusion as a search to receive and build on an effective recipe. Recipients socially proximate to the source of the knowledge have superior, though still imperfect, access to the original recipe. This advantage in access translates into higher fidelity reproduction that benefits the actor most significantly when the ingredients of the recipe display moderate interdependence. Simple recipes spread through the social network thoroughly, placing recipients both near and far on equal footing. Highly intricate recipes resist diffusion to even nearby actors. But for recipes of intermediate interdependence, nearby actors receive enough guidance from the template that local search delivers them an effective replica of the original knowledge on which they can build, while distant actors begin their search processes from such flawed starting points that subsequent efforts to receive and build on the interdependent recipe tend to fail.

4. Empirical corroboration

To test our hypothesis, we analyzed prior art citations to all US utility patents granted in May and June of 1990 (n = 17,264).[5] The data came from the Micro Patent database and NBER public access data on patents (Hall et al., 2001). Following much previous research, we view a prior art citation as evidence of knowledge diffusion: the applicant has successfully assimilated the knowledge underlying the original patent to a new setting and built on it. Our statistical approach is to estimate the likelihood that a focal patent receives a citation from a future patent as a function of several factors: the interdependence of the knowledge underlying the focal patent, the proximity of the inventors of the focal and citing patent in a social network, the interaction of interdependence and social proximity, and a set of control variables. The results of the estimation allow us to examine how the likelihood of citation by a socially proximate inventor compares to the likelihood of citation by a distant inventor as a function of knowledge interdependence. The crucial test of our hypothesis is whether the gap between the two probabilities peaks when the focal patent embodies moderately interdependent knowledge.

Patents and the meaning of citations

Patents and their citation patterns provide an attractive test bed for our hypothesis for several reasons. First, these citations have been carefully assigned. The US Patent Office requires all applicants to demonstrate awareness of their invention's precedents by citing similar 'prior art' patents. Patent examiners in each technological domain review and supplement the prior art references to ensure accurate and comprehensive citations. Second, consistent with our ontology of knowledge, technology historians have demonstrated that one can conceptualize patented inventions as combinations of preexisting technological components (Basalla, 1988). The process of invention therefore involves both the replication of prior discoveries and the extension of those discoveries to new applications and in new combinations. When a citation to prior art emerges on a new patent, it suggests that the inventor has both successfully received and built on the knowledge underlying the earlier patent. Third, Fleming and Sorenson (2001) have developed a technique for measuring the interdependence among the components of an invention. The technique draws on information uniquely available for patents and potentially difficult to duplicate in other settings.

This setting nevertheless also has its limitations. First, our analysis rests on the assumption that some potential knowledge recipients have better access to the template than others. If every patent fully revealed the inventor's underlying knowledge of the invention, this assumption would not hold. Inventor's incentives, however, minimize the likelihood of this problem. Patent applicants prefer to disclose as little as possible to limit their competitors' ability to benefit from their disclosure (Lim, 2001). Indeed, conversations with the US Patent Office indicate that applicants often intentionally obfuscate their descriptions to diminish the value of the knowledge revealed (Stern, 2001).

Second, the use of citations as an indicator of knowledge flows has been cast into doubt recently by the work of Alcacer and Gittelman (2004), who find that examiners add 40 percent of the citations found on US patents. On the one hand, this finding is comforting as it suggests that examiners actively work to prevent applicants from excluding citations to relevant prior art for strategic reasons, such as those mentioned above. It is nonetheless potentially problematic for our study to the extent that examiners most frequently insert socially proximate citations to patents of intermediate interdependence. The few studies that analyze the characteristics of examiner-added citations, however, show no evidence of such a bias (Alcacer and Gittelman, 2004; Sampat, 2004). Indeed, *self*-citations – which almost certainly reflect true knowledge flows – as frequently come from examiners as from inventors. This suggests to us that, on balance, examiner intervention *improves* the quality of patent data for our purposes and cannot account for our results. Consistent with this conclusion, Duguet and MacGarvie (2005) find that firms' patent citation patterns match their survey responses regarding technology acquisition and dispersion. At worst, if examiners add citations that do not reflect true knowledge flows and do so in an unbiased way, this should only add noise, increasing the difficulty of finding statistical support for our hypothesis.

Third, patents admittedly offer imperfect measures of invention. Inventors may limit their patent applications to a subset of their discoveries, and one must ask whether this selection process biases our results. Inventors most likely seek legal protection when a patent raises a meaningful barrier to imitation (e.g. when inventing around the patent proves difficult), when the invention will not quickly become obsolete, and when few alternative 'natural' defenses protect the knowledge (Levin et al., 1987). Of these conditions, the last seems most germane to our study. It implies that our sample may under-represent inventions that involve highly tacit, causally ambiguous and complex knowledge. Empirical research, however, suggests that this selection bias may not exist: Cohen et al. (2000), for example, find that firms in industries with complex products disproportionately choose to patent.

Finally, we recognize that patents represent but one embodiment of knowledge. Though we have no reasons to expect a priori that they should differ from other pieces of knowledge, they may. Despite this potential limitation on the scope of the applicability of our results, patents offer an excellent first test bed for our ideas for the reasons noted above.

Case-control design

Our unit of analysis is a patent dyad, one patent issued in May or June of 1990 and one issued later that may or may not cite the first. Hence our approach conceptually follows that of other studies of the likelihood of tie formation – in this case, the likelihood that a future patent builds on the knowledge embodied in one of our focal patents. These studies have typically estimated tie formation on the entire matrix of possible relations (e.g. Gulati, 1995; Podolny, 1994). This approach has two disadvantages. With large numbers of nodes, in this case patents, it can generate enormous, sparse matrices, increasing the difficulty of estimation and variable construction. In our situation, this method would generate nearly 20 billion dyads with only around 60,000 realized citations. In addition, this approach raises questions regarding network autocorrelation and the non-independence of repeated observations on the same patents across multiple observations in the error structure.

Instead, our analysis follows Sorenson and Stuart (2001) in adopting a case-control approach to analyzing the formation of ties (see Sorenson and Fleming, 2004, for an earlier application to patents). The case-control sampling procedure works as follows. We begin by including all cases of future patents, from July 1990 to June 1996, that cite any of our 17,268 focal patents: 60,999 in total. Since these citations occur, the dependent variable $Cite_{ij}$ takes a value of '1' for these cases to denote a realized citation. In addition, we pair each focal patent with four future patents that do not cite it (but that could have).[6] We set $Cite_{ij}$ to zero for these control cases. Though this generates a data set of 130,055 dyads, our analysis restricts the sample used for estimation to the 72,801 cases where both inventors reside in the US.[7] To address the fact that focal patents enter the data more than once, we report robust standard errors estimated without the assumption of independence across repeated observations of the same focal patent.

The use of a matched sample introduces one new problem. Logistic regression can yield biased estimates when the proportion of positive outcomes in the sample does not match the proportion of citations in the population (Prentice and Pyke, 1979; Scott and Wild, 1997). In particular, uncorrected logistic regression using a matched sample tends to produce underestimates of the factors that predict a positive outcome (King and Zeng, 2001). Large samples do not necessarily alleviate this problem.

We adjust the coefficient estimates using the method proposed by King and Zeng (2001) for the logistic regression of rare events (cf. Manski and Lerman, 1977). The traditional logistic regression model considers the dichotomous outcome variable a Bernoulli probability function that takes a value 1 with the probability π:

$$\pi_i = \frac{1}{1 + e^{-X\beta}},$$

where X represents a vector of covariates and β denotes a vector of parameters. Researchers typically use maximum likelihood methods to estimate β. King and Zeng (2001) prove that the following weighted least squares expression estimates the bias in β generated by oversampling rare events:

$$\text{bias}(\hat{\beta}) = (X'WX)^{-1}X'W\xi,$$

where $\xi = 0.5\,Q_{ii}[(1 + w_1)\hat{\pi}_i - w_1]$, the Q are the diagonal elements of $Q = X(X'WX)^{-1}X'$, $W = \text{diag}\{\hat{\pi}_i(1 - \hat{\pi}_i)w_i\}$, and w_1 represents the fraction of ones (citations) in the sample relative to the fraction in the population. At an intuitive level, one regresses the independent variables on the residuals using W as the weighting factor. Tomz (1999) implements this method in the relogit Stata procedure.

This case-control approach offers two principal advantages over the count models employed in most patent research. First, this method permits far more fine-grained controls for heterogeneity in citing patents. Count models preclude the possibility of controlling for detailed features of a citing patent. The ability to account for the attributes of the potential citing patents proves critical, however, to testing our hypotheses, which suggest that the ability of future inventors to receive and build on the original knowledge varies as a function of their social proximity. Second, analyzing citations at the level of the citing-patent/cited-patent dyad avoids the potential for aggregation bias inherent in count models.

Interdependence

Following Fleming and Sorenson (2001), we measure the complexity of the knowledge in a patent by observing the historical difficulty of recombining the elements that constitute it. Though it involves intensive calculation, the intuition behind the metric is straightforward: a technology whose components have, in the past, been mixed and matched readily with a wide variety of other components has exhibited few sensitive interdependencies. The measure considers the subclasses identified in a patent as proxies for the underlying components. Though in many cases subclasses correspond to identifiable physical components (such as in the example below), they do not always align so closely. Our measure, however, requires only that these subclasses define pieces of knowledge rather than physical components. Combining some pieces that interact sensitively to each other proves more difficult than connecting relatively independent chunks of knowledge.

We calculate the measure of interdependence, **k**, in two stages.[8] Equation 15.1 details our measurement of the ease of recombination – the inverse of interdependence – for subclass *i* used in patent *j*. We first identified every use of the subclass *i* in previous patents from 1980 to 1990.[9] The sum of the number of prior uses provided the denominator. For the numerator, we counted the number of different subclasses appearing with subclass *i* on previous patents. Hence, our measure increases as a particular subclass combines with a wider variety of technologies, controlling for the total number of applications, and captures the ease of combining a particular technology. To create our measure of interdependence for an entire patent, we averaged the inverted ease of recombination scores for the subclasses to which it belongs (equation 15.2).

$$\text{Ease of recombination of subclass } i \equiv E_i = \frac{\text{Count of subclasses previously combined with subclass } i}{\text{Count of previous patents in subclass } i} \tag{15.1}$$

$$\text{Interdependence of patent } j \equiv k_j = \frac{\text{Count of subclasses on patent } j}{\sum_{i \in j} E_i} \tag{15.2}$$

Intuitively, the measure operates as follows. Suppose a patent embodies subclasses that have been combined with a wide variety of subclasses, even in a handful of previous patents. This indicates that the patent's components do not have delicate interdependencies that prevent widespread recombination and the components can be mixed and matched independently. Such a patent receives a low value of **k**. Suppose instead that a patent embodies subclasses that have been combined, again and again, with the same small set of other subclasses. We presume those subclasses to be highly interdependent; their repeated joint appearance in patents suggests that the presence of one requires the appearance of the others. Hence the patent's **k** is high.

In addition to the measure's face validity, it has been validated externally via a survey

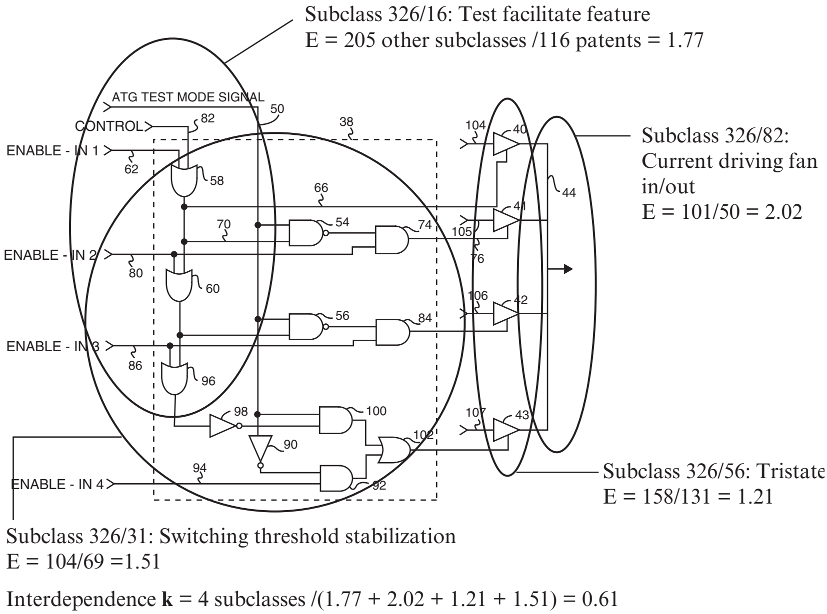


Figure 15.1 Calculation of interdependence for patent #5,136,185

of inventors. Fleming and Sorenson (2004) asked a sample of patent holders the following question, based on Ulrich's (1995) definition of interdependence: 'Modules are said to be coupled when a change made to one module requires a change to the other module(s) in order for the overall invention to work correctly. How coupled were the modules of your invention?' They then compared survey responses to calculated **k** for the corresponding patents and found a significant correlation between inventors' perceptions of coupling and the calculated degree of interdependence.

Concrete examples may clarify the metric further and help to link it to our core hypothesis. Consider a digital technology patent, #5,136,185, filed by the third author of this chapter. Figure 15.1 outlines the calculation of **k** for this patent and the mapping of the USPTO classification scheme to the components used. Here, 326/16 identifies the 'Test facilitate feature' subclass, which implements a testing mode within a semiconductor chip. Prior to its appearance here, this subclass had been recombined 116 times with 205 other components, implying an observed ease of recombination score of 205/116 = 1.77. Also, 326/56 indicates the 'Tristate' subclass, 326/82 points to 'Current driving fan in/out', and 326/31 identifies the 'Switching threshold stabilization' subclass (essentially a priority encoder). Figure 15.1 illustrates the location of these components on the circuit, the calculation of their ease of recombination scores, and the calculation of the patent's interdependence, **k** (= 0.61) – a level slightly above the mean **k** for our sample.

The invention described above assists engineers in testing the logic gates on new chips – a difficult task when chips can contain hundreds of thousands or even millions of such gates. Even though the patent appears to disclose much of the important information, it does not reveal the proprietary test generation algorithm, and how that algorithm manipulated the components (in particular, the 'test facilitate feature'). Without access to, or an understanding of, that algorithm, rivals could see the components of the knowledge in the patent but not how the components worked together. As a result, competitors faced an uphill battle in exploiting the knowledge. Even within the firm, effective transmission required the inventor to travel around the country to teach others how to use the technology. Similarly, competitors found it difficult to reproduce IBM's copper interconnect technology – another invention of intermediate complexity – until enough engineers defected to rivals to diffuse the relevant knowledge of how to fabricate the copper interconnect without contaminating the wafer's other materials (Lim, 2009).

By comparison, inventions involving extremely high levels of interdependence defy diffusion even within a social boundary. Plasmid preparation, for example, a biological technique, involves an intricately intertwined sequence of actions involving various chemicals, reagents and manual operations. As Jordan and Lynch (1992, p. 84) note, 'Although the plasma prep is far from controversial and is commonly referenced as a well established and indispensable technique, how exactly it is done is not effectively communicated, either by print, word of mouth, or demonstration.' On the other hand, inventions involving a low degree of interdependence diffuse rapidly. For instance, patent #4,927,016, one of the patents in the bottom quartile of the **k** range, involves the production of monoclonal antibodies. The industry associated with this technology has essentially become a commodity business since one can easily acquire all the necessary knowledge components by reading a textbook and piece them together without concern for sensitive interdependencies. Polymerase chain reaction, a technique for amplifying DNA sequences, has followed a similar route. Or, one might think of Sun's workstation technology. The modular design of its system has allowed rivals to match the performance of its hardware quickly, limiting the company's ability to maintain an advantage in the hardware market.

Social proximity

The analyses investigate the effect of knowledge complexity on the diffusion of knowledge to individuals whose close social connections to the source of knowledge give them better access to the template than individuals with distant or no connections have. For each of our 72,801 patent dyads, we develop one direct and two indirect indicators of social proximity between the inventors of the two patents in the dyad.

Proximity in a collaboration network Our most direct indicator measures the distance between inventors in a network of patent collaborators. The idea underlying this indicator is that an inventor gains access to a template via collaborators, collaborators of collaborators, collaborators of collaborators' collaborators, and so forth. Closer connections grant better access. To measure collaborative proximity, we use the methods and data of Singh (2005).[10] Consider the dyad consisting of patent i issued in May or June of 1990 and patent j issued at a later time t (before 1996). To compute the distance between i and j, Singh first constructs a network with a node for each discrete inventor who has been listed on any patent from 1975 until time t. An edge connects two inventors if they have collaborated on a patent during that period. The collaborative distance of a patent dyad is then the minimum number of intermediaries required to connect a member of the team of inventors listed on patent i to a member of patent j's team. If the two teams share a member, for instance, the distance is zero. If the teams have no common members but an individual listed on neither patent has collaborated with members of both *i*'s and *j*'s teams, the distance is one, and so forth. If no path connects members of the two teams, the distance is ∞. See Singh (2005) for a complete description of his approach.

Based on the distance measure, we construct three indicator variables for each dyad:[11]

Close collaboration$_{ij}$ = 1 if the distance between patents *i* and *j* is less than four; 0 otherwise.

Far collaboration$_{ij}$ = 1 if the distance between *i* and *j* is four or greater but less than ∞; 0 otherwise.

Unconnected$_{ij}$ = 1 if no path connects *i* and *j*.

The shorter the path between *i* and *j*, the better the access to the template enjoyed by the team involved in patent *j*. Our core hypothesis is that this superior access translates into a higher probability of citation especially when the components of patent *i* display intermediate interdependence. Accordingly, we expect the gap in citation probability between a close and a far inventor – the probability that a close inventor cites a focal patent minus the probability that a far inventor cites the patent – to peak at an intermediate level of $\mathbf{k}$.

Although our collaborative distance measure provides direct evidence of access and we believe that it captures many of the important connections between inventors, inventors also have many other types of relations that might also facilitate access. For example, a potential recipient might be a friend of the source even if they have never collaborated. Attempting to identify all of the potential relationships existing in any population of individuals is not feasible, but we can examine two factors – geographic proximity and joint organizational membership – that tend to structure social relationships and therefore may proxy for unobserved social paths between our source-recipient dyads. As McPherson et al. (1992, p. 154) note: 'Homophily structures the flow of information and other social resources through the network so that the dimensions themselves stand as proxies for the number of intervening steps in transmissions through the system.'

Geographic proximity Space represents one important dimension that structures social interaction. Indeed, some of the earliest literature on social networks emphasized the dramatic decline in the likelihood of a social relation as two parties became increasingly distant (Bossard, 1932; Park, 1926). Accordingly, we develop a measure of geographic proximity for each patent dyad:

Geographic proximity$_{ij}$ = the natural log of the distance in miles between the first inventors listed on patents *i* and *j* multiplied by negative one (so that larger values indicate greater proximity).[12]

As with our direct measure of social proximity, we expect geographic proximity to have the greatest impact on citation likelihood when the potentially cited patent displays moderate interdependence.

Organizational proximity Social networks also concentrate within foci, such as organizations (Feld, 1981). On a daily basis, most fully employed individuals spend more waking hours engaged in work than in any other activity. Employees regularly meet other employees through work to cooperate on projects, to confer on decisions, to transfer information, and to socialize. Hence, we use employment at the same patent assignee as another indicator of social proximity:

$Organizational\ proximity_{ij}$ = 1 if the same organization owns both patents in a dyad, 0 otherwise.

We expect common ownership to boost citation likelihood, especially for focal patents of moderate interdependence.

We test our hypothesis by regressing $Cite_{ij}$ on the indicators of social proximity directly, the indicators interacted with $\mathbf{k}$, and the indicators interacted with $\mathbf{k}^2$. We expect social proximity to boost citation probability directly. The core test of our hypothesis resides not in the direct effects but in the interaction terms: the impact of proximity on citation probability should have an inverted-U relationship with respect to interdependence $\mathbf{k}$.[13]

In light of our empirical context, patent citations, it is useful to elaborate our expectations about the direct effect of $\mathbf{k}$ on citation likelihood. Our hypothesis describes the impact of interdependence on the *gap* between near and distant actors' success in receiving and building on knowledge. We examine this gap by examining *interactions* of $\mathbf{k}$ and $\mathbf{k}^2$ with social distance. In developing the hypothesis, however, we also paint a picture of the *direct* impact of $\mathbf{k}$ on knowledge reproduction: we suggest that greater interdependence increases the difficulty for a party of receiving and building on prior knowledge, regardless of the party's distance from the source. This argument concerns an actor's success in receiving and building on knowledge *conditional on an attempt to do so being undertaken*. Patent citation data nevertheless reflect not only success conditional on an attempt being undertaken, but also the sheer number of attempts being undertaken. We have reason to believe that the number of attempts may rise with interdependence, simply because interdependence increases the fertility that comes from mixing and matching components (Fleming and Sorenson, 2001). Accordingly, we offer no hypothesis about the direct effects of $\mathbf{k}$ on citation rates. Instead, we focus on the *gap* between near and distant actors' citation rates, which should have a robust inverted-U relation to interdependence. (See the Appendix for a more detailed treatment of this point.)

Controls

The non-monotonic interactions between interdependence and proximity that we predict – if found in the data – lend themselves to few alternative interpretations. The models nevertheless include as controls several of the most important variables used in prior patent studies (e.g. Lanjouw and Schankerman, 2004).

Activity control The activity control accounts for the typical number of citations received by a patent in the same technological areas as the focal patent. In a first stage, we calculated the average number of citations that each patent in a particular USPTO class received from patents granted between January of 1985 and June of 1990 (equation

15.3).[14] We then weighted these parameters according to the patent's class assignments (equation 15.4), where p_{ik} indicates the proportion of patent k's subclass memberships that fall in class i.

$$\text{Average citations in patent class } i \equiv \mu_i = \frac{\sum_{j \in i} \text{Citations}_j \text{ (before 7/90)}}{\text{Count of patents } j \text{ in subclass } i} \quad (15.3)$$

$$\text{Technology mean control patent } k \equiv M_k = p_{ik}\mu_i \quad (15.4)$$

The models also include controls for several other factors. *Same class* is a dummy variable denoting whether the two patents in each dyad belong to the same primary technological class. *Recent technology* is the mean of the patent numbers of the focal patent's prior art (higher numbers indicating more recent technology).[15] The models include counts of two types of backward patent citation. First, they include a tally of the number of citations to patent *prior art*. Second, the models include a control for the number of *non-patent prior art* citations (e.g. references to published articles). *Number of classes* is a count of the number of major classes and *number of subclasses* is a count of the number of subclasses to which the focal patent is assigned. Descriptive statistics appear in Table 15.1.[16]

5. Results

The results appear in Table 15.2. Model 1 estimates the effects of the control variables alone, and model 2 introduces interdependence, **k**.

Model 3 provides the first test of our core hypothesis by interacting interdependence with collaboration-based indicators of social proximity. The results provide three pieces of support for the hypothesis. First, the positive sign on **k** x *Close collaboration* coupled with the negative sign on $\mathbf{k}^2$ x *Close collaboration* indicates that the gap in citation probability between close and unconnected inventors rises and then falls, peaking when the source knowledge displays moderate interdependence. (Recall that *Unconnected* is the excluded category, so the coefficients related to *Close collaboration* capture differences between close and unconnected inventors.) Second, by subtracting the coefficients for *Far collaboration* from the coefficients for *Close collaboration*, we see that the largest gap between close and far inventors also appears for moderate **k**. Third, the coefficient estimates suggest that the greatest gap between far and unconnected inventors arises for moderate **k** (though with much smaller magnitude; see below). In sum, our primary measure for social proximity provides strong support for our core hypothesis.[17]

Model 4 adds interactions of interdependence with geographic and organizational proximity. Both proxies for social proximity display the expected inverted-U relationship, though only the results for geographic proximity show strong statistical significance. Coefficients for the collaboration-based measures retain their signs and significance, as do most of the coefficients for the control variables.

Based on model 4, Figure 15.2 traces out, as a function of interdependence, how many times more likely a citation is for collaboratively close pairs of inventors than for unconnected pairs, for close pairs than for far pairs, and for far pairs than for unconnected pairs. (We set all other variables to their mean values for the purpose of creating this chart.) The figure shows vividly that the maximal difference in citation probabilities

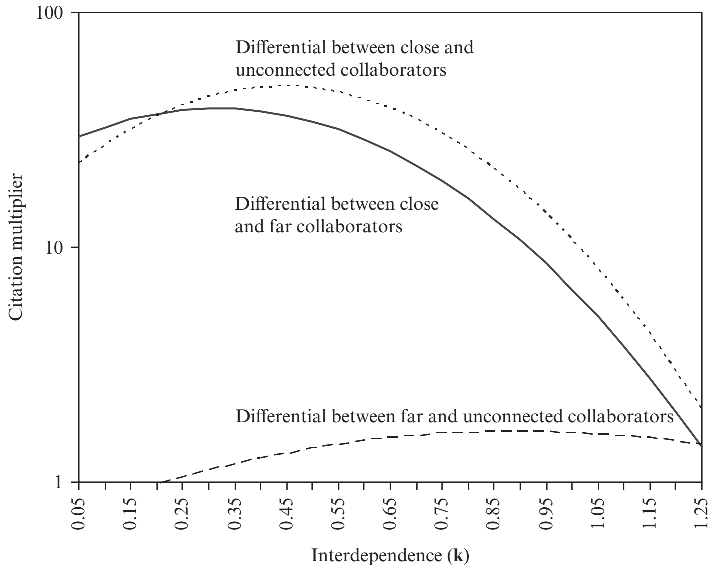
Table 15.1 Descriptive statistics and correlations

	Mean	SD	2	3	4	5	6	7	8	9	10	11	12
1. **k**	0.49	0.30	0.03	−0.02	−0.07	.00	−.11	.10	.07	−.03	−.05	−.29	−.35
2. *Close collaboration*	0.07	0.25		−0.21	0.14	0.30	0.17	0.00	0.08	0.01	0.01	−0.01	0.00
3. *Far collaboration*	0.23	0.42			−0.06	0.01	0.01	0.12	0.15	−0.01	0.03	0.02	0.05
4. *Organizational proximity*	0.10	0.33				−0.09	−0.03	−0.06	−0.07	0.02	−0.04	−0.02	−0.03
5. *Geographic proximity*	−6.50	1.96					0.20	0.00	0.05	0.05	0.01	0.00	0.01
6. Same class	0.26	0.44						0.12	0.08	0.04	0.02	−0.11	−0.01
7. Activity control	1.25	0.42							0.42	0.01	0.06	−0.02	0.08
8. Recent technology	3.97	0.62								−0.14	0.09	0.05	0.09
9. Backward patent citations	9.83	8.88									0.13	0.07	0.12
10. Backward non-patent citations	1.46	4.24										0.06	0.10
11. Number of classes	1.85	0.97											0.49
12. Number of subclasses	4.53	3.43											

*Table 15.2 Rare events logit models of the likelihood of a focal patent receiving a citation from a future patent**

	Model 1	Model 2	Model 3	Model 4
k		0.863•••	−1.599•	−1.305••
		(0.257)	(0.644)	(0.378)
$\mathbf{k}^2$		−0.203•••	1.116•••	1.051•••
		(0.086)	(0.209)	(0.156)
k × *Close collaboration*			3.242•	4.327••
			(1.670)	(1.659)
$\mathbf{k}^2$ × *Close collaboration*			−3.428•••	−4.881•••
			(0.708)	(0.725)
k × *Far collaboration*			1.569••	1.899••
			(0.573)	(0.627)
$\mathbf{k}^2$ × *Far collaboration*			−0.802•••	−1.056•••
			(0.162)	(0.262)
k × *Geographic proximity*				0.325•••
				(0.078)
$\mathbf{k}^2$ × *Geographic proximity*				−0.241•••
				(0.031)
k × *Organizational proximity*				0.547
				(0.679)
$\mathbf{k}^2$ × *Organizational proximity*				−0.508•
				(0.232)
Close collaboration	3.952•••	3.979•••	3.660•••	2.925••
	(0.628)	(0.618)	(1.148)	(1.135)
Far collaboration	0.224•	0.249••	0.244••	−0.359
	(0.090)	(0.089)	(0.089)	(0.246)
Geographical proximity	0.041•••	0.041•••	0.045•••	0.053•••
	(0.012)	(0.012)	(0.011)	(0.012)
Organizational proximity	0.457•••	0.431•••	0.423•••	0.292
	(0.118)	(0.119)	(0.116)	(0.355)
Same class	4.800•••	4.820•••	4.797•••	4.784•••
	(0.084)	(0.085)	(0.083)	(0.083)
Activity control	0.503•••	0.515•••	0.469•••	0.481•••
	(0.097)	(0.098)	(0.095)	(0.096)
Recent technology	0.268	0.278	0.226	0.245
	(0.147)	(0.144)	(0.161)	(0.147)
Backward patent citations	0.022•••	0.021•••	0.020•••	0.021•••
	(0.005)	(0.005)	(0.005)	(0.005)
Backward non-patent citations	0.011	0.014	0.010	0.010
	(0.009)	(0.009)	(0.008)	(0.009)
Number of classes	0.184•••	0.209•••	0.204•••	0.201•••
	(0.047)	(0.048)	(0.047)	(0.046)
Number of subclasses	−0.030•	−0.016	−0.017	−0.023
	(0.013)	(0.014)	(0.013)	(0.013)
Constant	−12.28•••	−12.89•••	−12.21•••	−11.91•••
	(0.586)	(0.657)	(0.746)	(0.697)
Log-likelihood	−33,772.4	−33,751.1	−33,738.2	−33720.2

Note: * 72,801 dyads (52% realized ties versus .0004% in population); • $p < 0.05$; •• $p < 0.01$; ••• $p < 0.001$



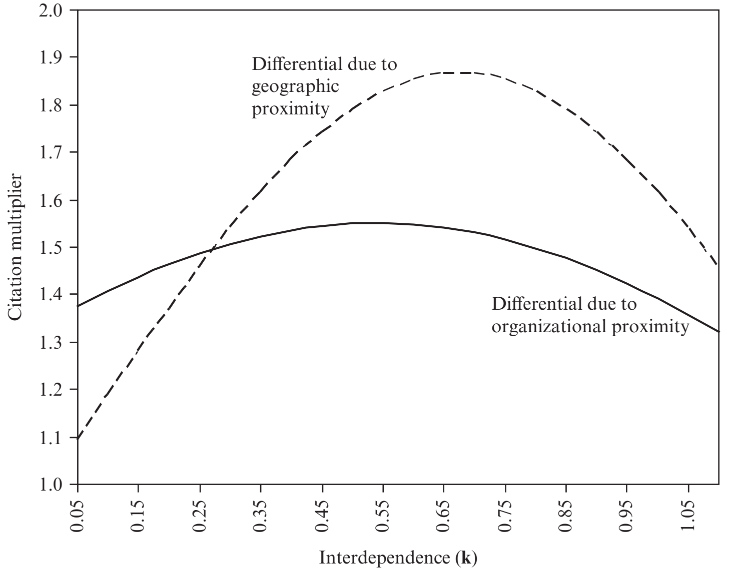
Note: The line labeled 'differential between close and unconnected collaborators' shows, as a function of **k**, how many times more likely a citation is in a dyad of patents whose inventors cannot reach one another through the collaboration network (path length < 4) relative to a dyad whose inventors are unconnected in the network. When **k** = 0.45, for instance, a citation is 48 times more likely. The other two lines provide the same information for pairs of actors who are close versus far (path length between 4 and ∞) in the collaboration network and for pairs of actors who are far versus unconnected. The figure is based on model 4 of Table 15.2 for inventors from different organizations, with all variables other than **k** and the collaboration network indicators set to their mean values.

Figure 15.2 Citation multiplier for proximate versus distant actors in the collaboration network as a function of interdependence

between close pairs and unconnected pairs arises when the focal patent displays moderate interdependence. The same is true of the difference between close and far pairs. Figure 15.2 also shows that the citation difference between far and unconnected inventors – while consistent with our hypothesis and statistically significant – is much, much smaller. This suggests that for access to knowledge, the value of a social connection to the source drops off rapidly with the number of intervening intermediaries, echoing the findings of Singh (2005).

Figure 15.3 shows, as a function of interdependence, how many times greater the probability of citation is between geographically proximate actors than between geographically distant actors. It does likewise for pairs of inventors in the same organization versus pairs in different organizations. In both cases, the benefits of social proximity rise and then fall with **k**, peaking when the source knowledge displays moderate interdependence. This provides graphical affirmation of our hypothesis. In both Figures 15.2 and 15.3, the peak differences fall within the range of actual **k** in our data – in fact, within one standard deviation above the mean.

In addition to being significant, the effects associated with our hypothesis can have substantial economic import. For source knowledge that is simple (**k** ~ 0), an inventor



Note: The line labeled 'differential due to geographic proximity' shows, as a function of **k**, how many times more likely a citation is in a dyad of patents when the inventors' addresses on the patents reside 10 miles apart than when they reside 3000 miles apart. When **k** = 0.65, for instance, the multiplier is 1.87 (i.e. 87% more likely). The line labeled 'differential due to organizational proximity' shows, as a function of **k**, how many times more likely a citation is in a dyad of patents when the same organization owns both patents relative to when they are owned by different organizations. The figure is based on model 4 of Table 15.2, with all variables other than **k** and geographic and organizational proximity set to their mean values.

Figure 15.3 Citation multiplier for proximate versus distant actors (in geography and organizational space) as a function of interdependence

close in the collaboration network is 30 times more likely than a far inventor to cite a focal patent. For knowledge of moderate interdependence at the gap-maximizing level of **k** shown in Figure 15.2, this number rises to 39 times. As knowledge becomes more complex, the number falls, becoming a mere 7 times at **k** = 1. For close and unconnected inventors, the figures are 23 times, 48 times, and 11 times, respectively.[18] Similarly, contrast an inventor 10 miles from the source of knowledge and another 3000 miles away (both collaboratively unconnected to the source and in different organizations). When **k** ~ 0, the first inventor is 9 percent more likely that the second to cite the source. When **k** is at the gap-maximizing level, the probability rises to 87 percent. It then falls to 61 percent for **k** = 1. Such differences in citation likelihood are far from negligible.

Despite the apparent consistency of our results with our expectations, proximity – collaborative, geographic, or organizational – may reflect factors other than the strength of social connections, factors that might also influence the quality of one's access to the template. Actors proximate to a given patent might, for instance, work on similar technical problems and therefore more readily absorb the knowledge embodied in the patent (Cohen and Levinthal, 1990). Any factor that improves access to the template should have the effect that we hypothesize. It is natural to interpret the proximity measures as indicators of social contact, as we do. It is difficult, however, to rule out all other factors that the proximity measures might reflect.

Similarly, our interdependence measure may capture not only the complexity of an item of knowledge but also its breadth of applicability. Our results might then reflect a process in which low-**k** knowledge is broadly applicable and diffuses widely; moderate-**k** knowledge is of particular interest to select groups who tend to be socially proximate to the inventor; and high-**k** knowledge is of such narrow application that it diffuses very narrowly. This would produce a pattern in which actors socially proximate to a source of knowledge most frequently receive and build on it if the knowledge has moderate **k**. The driving force under this alternative interpretation is not the relative ability of different actors to search in the face of complexity but the relative interest that different actors have in obtaining knowledge. The alternative interpretation raises the question of precisely what makes an item of knowledge broadly or narrowly applicable. Knowledge becomes broadly applicable in part because it is modular and therefore can mix and match with other pieces of knowledge across a wide range of circumstances. Applicability, then, may capture the interdependence of a piece of knowledge (especially if one defines interdependence broadly and not in a narrow technological sense). To the extent that applicability reflects interdependence, we return to our original core hypothesis: individuals proximate to the source of some knowledge have the greatest advantage in receiving and building on knowledge of moderate interdependence/applicability.

6. Discussion

The analysis of patent citation patterns supports our basic theoretical perspective on knowledge diffusion: search in the space of possible combinations of ingredients offers a useful lens for understanding the flow of knowledge. Recipients socially proximate to the source of the knowledge have preferential access to the original success, which serves as a template during efforts to receive and build on the knowledge. All recipients, socially near and far, compete on equal footing when receiving and extending simple knowledge; incremental search suffices to reproduce simple knowledge, so guidance from a prior success has little value. Highly complex knowledge, on the other hand, equally resists diffusion to both classes of would-be recipients. Hence, at both extremes of complexity, the close recipient has no lasting advantage over the distant. In contrast, for knowledge whose ingredients display a moderate degree of interdependence, superior but imperfect access to the template translates into greater success in receiving and building on pre-existing knowledge. The close recipient can complete its initially imperfect replica via local search, but local search alone cannot guide the distant recipient to an accurate replica. Thus in our patent data, the largest gap between the ability of a close recipient to receive and build on prior knowledge relative to the ability of a distant recipient arises when the cited patent involves moderate interdependence. This result appears when social distance is measured by proximity in a collaboration network as well as when geographic and – to a lesser extent – organizational proximity proxy for social distance.

Our findings have an array of practical and theoretical implications, especially for the issue of knowledge inequality across social borders. Consider the graph of a typical social network. It is quite common in such a graph to observe patches of actors with dense connections amongst themselves and areas of sparse connections between patches

(Owen-Smith and Powell, 2004). The dense patches may reflect firms, for instance, or geographic regions. Actors within each patch sit socially proximate to one another but relatively distant from actors in other patches. A question of great practical importance is: When does knowledge diffuse within the patch where it originated but not across the thin areas into other patches? When will knowledge diffuse within a firm but not to competitors, or within a region but not to other locales? When is inequality of knowledge sharpest across social borders? Our results suggest that the nature of the knowledge, specifically its degree of complexity, plays a critical role. One might initially suspect that highly complex knowledge, the most difficult to reproduce, would create the greatest inequality across boundaries. Yet this intuition ignores the fact that inequality in its sharpest form requires *some* diffusion: to create the most inequity across social boundaries, knowledge must creep up to the edge of the thick patch of connections in which it originated but not beyond. This phenomenon, we have argued, most likely occurs for moderately complex knowledge.

Accordingly, the results suggest a resolution to the replication/imitation dilemma that has puzzled evolutionary economists and strategy scholars. To achieve a competitive advantage from knowledge, a firm must typically leverage that knowledge across multiple applications, for example, across all its production facilities (Winter, 1995). Yet any would-be replicator with a valuable piece of knowledge faces a dilemma: the profits produced by its original knowledge attract the envious attention of imitators. Valuable knowledge provides a source of *sustained* advantage only to the extent that it lends itself to replication yet defies imitation. Unfortunately for the innovator, replication and imitation typically go hand in hand (Nelson and Winter, 1982). Our results suggest, however, that replication-without-imitation is especially likely when the target knowledge entails moderate complexity. This micro-level phenomenon may manifest itself in outcomes at the industry level. One might expect that, *ceteris paribus*, industries based on moderately complex knowledge will display especially wide intra-industry dispersion in long-run financial returns. We leave this promising hypothesis for future research.

The results also speak interestingly to the literature on the geographic agglomeration of industries. Researchers frequently cite knowledge spillovers as a prominent reason that firms within an industry cluster together (Krugman, 1991; Marshall, 1890) and congregate near universities (Zucker et al., 1997). Our results certainly support this point of view: dense social networks, which tend to localize geographically, give firms and individuals close to the source of knowledge an important advantage in reproducing and building on the knowledge. This begs the further question, why do some industries cluster while others do not? Though research on economic geography points out that knowledge spillovers can contribute to agglomeration, it does not identify *what type of knowledge* most likely engenders these clusters. Our findings suggest that industries that rely on moderately complex knowledge more commonly form industrial districts (cf. Sorenson, 2004). Simple knowledge can diffuse far and wide because incremental search efforts can substitute for high-fidelity communication. As the complexity of knowledge increases, a gap emerges between local diffusion and distant diffusion; thus, the potential return to locating near to innovators rises.

In addition to influencing geographic agglomeration and industry structure, the nature of the underlying knowledge used by a firm may have implications for organizational design. Firms have both formal and informal structures that influence the degree to which actors within the firm interact with each other. Managers can influence who likely interacts with whom through the assignment of individuals to facilities, the design of laboratories and factories, and the structure of reporting relationships (Allen, 1977). To distribute knowledge effectively, a firm might usefully expend resources to foment close and dense social connections between sources and intended recipients of complex knowledge, while letting networks remain sparse elsewhere. Indeed, leaders might fruitfully construe the task of knowledge management *not* as the construction of central databases of information (as sometimes presented today), but rather as an effort to build social networks that match the nature and intended flow of knowledge. Effective organizational design, however, surely requires a deeper understanding of how social structure affects knowledge diffusion than considered here; networks have subtle features and nuances that doubtlessly influence their ability to convey knowledge, both simple and complex (Hansen, 1999).

To this point, our argument has assumed that the degree of interdependence between combinations of components remains fixed. In the long term, however, the effective interdependence of knowledge may change. Firms and inventors can invest in R&D to specify interfaces and embed knowledge within physical components, thereby reducing the difficulty of combining a particular combination of components with other elements in the future (Baldwin and Clark, 2000). In structuring knowledge, managers must perform a delicate balancing act. Isolating interdependencies within substructures has important attractions, including the ability to perform a greater number of independent experiments (Baldwin and Clark, 2000) and the capacity to adjust more readily to environmental shifts (Levinthal, 1997). Engineering curricula support this preference with a strong emphasis on reliability, black box design techniques, and the re-use of previously combined components (e.g. Mead and Conway, 1980). Such modularization, however, also entails frequently overlooked costs. Designing and implementing an architecture that isolates interdependencies within substructures involves considerable engineering costs (O'Sullivan, 2001). But those direct costs potentially pale in comparison to the indirect costs – the opportunities that the lack of complexity opens for new entrants (Rivkin, 2000), the reduction in variety from which developers can select (Christensen et al., 2002), and the constraints on potential performance (Fleming and Sorenson, 2001). Managers who manipulate interdependencies should recognize that they simultaneously alter the propensity of knowledge to flow to actors near *and* far.

Despite the costs of modularizing, a secular trend towards modularization may influence the evolution of industries, creating a distinctive pattern. Direct costs likely strike firms as more tangible than indirect costs as they decide where to direct R&D effort. Thus, firms may over-invest in less complex technology as they seek to maximize efficiency. As this process reduces the effective interdependence of the knowledge being diffused, knowledge should flow more easily, generating two industry-level patterns. First, an industry that begins its life in a concentrated region should become less concentrated geographically as the advantage of preferential access to the template declines (for related ideas, see Audretsch and Feldman, 1996; Stuart and Sorenson, 2003).[19] Second, the move towards less complex knowledge likely reduces differentiation across firms' products over time, leading to more intense price competition and efforts to control standard interfaces and key modules – a pattern identified in the product lifecycle literature.

To reiterate, our results demonstrate that knowledge complexity importantly influences the dynamics of diffusion. Specifically, a socially proximate actor's advantage over a distant actor in obtaining and building on knowledge peaks when the components underlying the knowledge display intermediate interdependence. Though our empirical results come from patent data alone, the basic logic of our hypotheses applies to knowledge in general, not just the knowledge underlying inventions. Hence, future research might usefully examine these dynamics across a wide range of applications – including organizational learning, the diffusion of management practices, knowledge management, and the sustainability of knowledge-based competitive advantage.

Notes

* All parties contributed equally to this research. We thank Jasjit Singh for generously sharing his inventor collaboration network data. We also appreciate the helpful comments of George Baker, Matt Bothner, Koen Frenken, Bob Gibbons, Jerry Green, Rebecca Henderson, Bruce Kogut, Dan Levinthal, Woody Powell, Rosemarie Ham Ziedonis, three anonymous reviewers, and seminar participants at Harvard, the University of Michigan, MIT, New York University, Ohio State University, the University of Toronto, Washington University, and Wharton. All errors remain our own. Harvard Business School's Division of Research provided financial support. This chapter was originally published in *Research Policy*, **35** (2006), 994–1017. Subsequent versions of this chapter have been presented at the annual conferences of the Academy of Management, the European Group on Organizational Studies, and the European Meeting on Applied Evolutionary Economics, as well as at the ESF Exploratory Workshop on Evolutionary Economic Geography.

1. Hansen (1999) also focuses on the interplay between social relations and knowledge flow. His research differs from ours in three respects: (1) it does not explore the issues related to recipient search as a mechanism for the interplay, (2) it focuses on the strength of the connection between inventors rather than social proximity in a network; and (3) it analyzes the effects of a *portfolio* of relations rather than the characteristics of a connection in a dyad. We nonetheless see close parallels that we revisit in the discussion section.
2. In this game, one child whispers a message into the ear of another, who then whispers what she heard into the ear of a third child and so forth. At the end, the final person announces the message he heard and the first person reveals the message that she originally whispered; the two usually differ dramatically.
3. This assumption limits the applicability of our theory to innovations that involve multiple components. This restriction should not severely constrain its scope, however; few innovations do not involve the combination of multiple physical components or processes. For example, even the synthesis of nylon, a polymer, involved the integration of several distinct processes (Smith and Hounshell, 1985).
4. Though not considered here, one might also consider the importance of tie 'strength'. Weak ties have long reach but low bandwidth; thus, they operate most prominently in the diffusion process when transferring only short, simple messages (Hansen, 1999).
5. We constructed this dataset in the course of prior research. For details on its construction, see Fleming and Sorenson (2001).
6. We chose four patents for the 'control' group so that the sample would have a roughly equal proportion of realized and unrealized dyads. Although some feel that conditioning on important factors improves the statistical power of a case-control sample (e.g. Jaffe et al., 1993, implicitly make such an argument in drawing controls from the same classes as the citing patents), the ideal method of selecting controls remains an open debate. Matching controls to cases on one or more dimensions can lead to two problems in particular that concern us. First, correcting the logit for over-sampling on the dependent variable requires that one knows the sampling probabilities (King and Zeng, 2001); matching controls to cases precludes the possibility of calculating this information. Second, matching on an endogenously determined factor risks generating biased results (e.g. when investigating diffusion processes, one would not want to consider the geographic distribution of activity exogenous). Given these concerns, we sample future patents at random and control for heterogeneity in the estimation.
7. Including the foreign inventors does not change the results qualitatively.
8. Our measure **k** is related to but distinct from the parameter K in the NK simulation models that have become popular in theoretical work on complex systems (Kauffman, 1993). In NK simulations, the contribution of each element in a system to overall system fitness depends on the states of K other elements. K is set by the modeler and, like our empirically measured **k**, reflects the degree of interdependence among components in a system. Despite the conceptual linkage between our measure **k** and Kauffman's K, we do not purport to have measured his K in a literal sense. For instance, our **k** does not equal the number of elements that affect the contribution of each focal element.

9. Some might worry about the stability of this measure over time. To test its robustness, we constructed a second **k** measure using data from 1790 to 1990. That measure yielded a qualitatively identical set of results.
10. Breschi and Lissoni (2002) independently developed an equivalent approach.
11. Though the magnitude of the gap shrinks, our results remain qualitatively robust to shifting the dividing line between close and far from a path length of three to a length of four. We use three categories rather than the distance measure itself for three reasons: (1) calculating the precise distance for the longer paths in these data would increase the time required to compute it by orders of magnitude (i.e. by months); (2) dummy variables for individual path lengths lead to some small cell sizes and concomitantly unstable coefficient estimates; and (3) given our interaction with a quadratic, we find the results of the categorical coding far easier to interpret and understand.
12. All patents list the home address of the inventor on the front page of the patent application. To locate each inventor, we match the inventor's 3-digit zip code to the latitude and longitude of the center of the area in which the inventor resides based on information from the US Postal Service. We then use spherical geometry to calculate the distance between the points. The USPTO includes 5-digit zip information, but we choose to reduce measurement error by using cleaned data. CHI, an information provider, has called every patent holder to verify the inventor's location; however, it records this information only at the 3-digit level.
13. We mean-deviate the variables before creating the interaction terms to facilitate interpretation of the effects (Friedrich 1982). For collaborative proximity, we use $Unconnected_{ij}$ as the excluded category.
14. We allow all patents issued between January 1985 and 30 June 1990 to enter the estimation of the activity control, meaning that the patents used to calculate it vary in the time during which they can receive citations. Alternatively, we could select a small set of patents from 1985 and base the measures on the subsequent five years of citations; however, this approach would ignore the patent activity just prior to our sample.
15. This variable made use of the fact that the USPTO assigns patent numbers sequentially. This assignment pattern generates a correlation between a patent number and the grant date of the patent of 0.98.
16. We also considered as a control variable the time between the issuance dates of the focal and potentially citing patents in each dyad. Exploratory analysis revealed small effect sizes (though typically significant), and inclusion of the time control had no meaningful impact on the coefficients of central interest.
17. Since the high correlation between a term and its square can force estimates to take opposing signs, we further tested the validity of our non-monotonic effect in two ways: (1) in unreported estimates (available from the first author), we re-estimated the models using a log-quadratic specification and found qualitatively identical results. Since this functional form can capture decreasing returns without a significant coefficient on the quadratic term, it is less sensitive to these problems; (2) we estimated a model with only the linear term and interactions and then entered the quadratic terms. In all cases, the addition of the quadratic terms significantly improved the model. (For example, in model 4, the addition of the quadratic **k** and its interactions has a $\chi^2 = 70.4$, significant at $p < .00001$ with five degrees of freedom.)
18. These figures assume that the two inventors are 665 miles from one another (the average distance in our sample) and work for different organizations.
19. This pattern seems consistent with the evolution of the software industry, for instance. Early on, knowledge localized to an extreme: understanding of a new piece of code resided in the head of a single developer or a small group of developers in a university, government, or large corporate computing facility. Inventors developed local languages for specific hardware. Over time, programmers developed techniques for reducing the interdependencies in code. Higher-level languages such as Cobol and C allowed programmers to divorce code from specific hardware. Meanwhile, software production has dispersed geographically – beyond Silicon Valley, Route 128, and IBM's Armonk home, to Seattle, Austin, and even Bangalore.

Bibliography

Alcacer, Juan and Michelle Gittelman (2004), 'How do I know what you know? Patent examiners and the generation of patent citations', *Review of Economics and Statistics*, **88**, 774–9.

Allen, Thomas J. (1977), *Managing the Flow of Technology: Technology Transfer and the Dissemination of Technological Information Within the R&D Organization*, Cambridge, MA: MIT Press.

Argote, Linda (1999), *Organizational Learning: Creating, Retaining and Transferring Knowledge*, Boston: Kluwer.

Arrow, Kenneth J. (1962), 'Economic welfare and the allocation of resources for invention', in Richard Nelson (ed.), *The Rate and Direction of Inventive Activity*, Princeton, NJ: Princeton University Press, pp. 609–24.

Arrow, Kenneth J. (1974), *The Limits of Organization*, New York: Norton.

Audretsch, David B. and Maryann P. Feldman (1996), 'Innovative clusters and the industry life-cycle', *Review of Industrial Organization*, **11**, 253–73.
Baker, Wayne E. and Robert R. Faulkner (2004), 'Social networks and loss of capital', *Social Networks*, **26**, 91–111.
Baldwin, Carliss Y. and Kim B. Clark (2000), *Design Rules: The Power of Modularity*, Cambridge, MA: MIT Press.
Basalla, George (1988), *The Evolution of Technology*, Cambridge: Cambridge University Press.
Boschma, Ron and Ron Martin (2007), 'Editorial: constructing an evolutionary economic geography', *Journal of Economic Geography*, **7**, 537–48.
Bossard, James S. (1932), 'Residential propinquity as a factor in marriage selection', *American Journal of Sociology*, **38**, 219–24.
Breschi, Stefano and Francesco Lissoni (2002), 'Mobility and social networks: localized knowledge spillovers revisited', Working Paper, Bocconi University.
Burt, Ronald S. (1987), 'Social contagion and innovation: cohesion versus structural equivalence', *American Journal of Sociology*, **92**, 1287–335.
Burt, Ronald. S. (1992), *Structural Holes: The Social Structure of Competition*, Cambridge, MA: Harvard University Press.
Chew, W. Bruce, Timothy Bresnahan and Kim Clark (1990), 'Measurement, coordination, and learning in a multiplant network', in Robert S. Kaplan (ed.), *Measures for Manufacturing Excellence*, Boston: Harvard Business School, pp. 129–62.
Christensen, Clayton M., Matt Verlinden and George Westerman (2002), 'Disruption, disintegration, and the dissipation of differentiability', *Industrial and Corporate Change*, **11**, 955–93.
Cohen, Wesley M. and Daniel Levinthal (1990), 'Absorptive capacity: a new perspective on learning and innovation', *Administrative Science Quarterly*, **35**, 128–52.
Cohen, Wesley M., Richard R. Nelson and John P. Walsh (2000), 'Protecting their intellectual assets: appropriability conditions and why U.S. manufacturing firms patent (or not)', W7552, National Bureau of Economic Research.
Coleman, James S., Elihu Katz and Herbert Menzel (1957), 'The diffusion of an innovation among physicians', *Sociometry*, **20**, 253–70.
Coleman, James S., Elihu Katz and Herbert Menzel (1966), *Medical Innovation: A Diffusion Study*, New York: Bobbs-Merrill.
Dahl, Michael and Olav Sorenson (2009), 'The embedded entrepreneur', *European Management Review*, **6**, 172–81.
Davis, Gerald F. and Henrich R. Greve (1997), 'Corporate elite networks and governance changes in the 1980s', *American Journal of Sociology*, **103**, 1–37.
Duguet, Emmanuel and Megan MacGarvie (2005), 'How well do patent citations measure flows of technology? Evidence from French innovation surveys', *Economics of Innovation and New Technology*, **14**, 375–93.
Durkheim, Emile (1912), *The Elementary Forms of Religious Life*, Oxford: Oxford University Press.
Feld, Scott L. (1981), 'The focused organization of social ties', *American Journal of Sociology*, **86**, 1015–35.
Fleming, Lee and Olav Sorenson (2001), 'Technology as a complex adaptive system: evidence from patent data', *Research Policy*, **30**, 1019–39.
Fleming, Lee and Olav Sorenson (2004), 'Science as a map in technological search', *Strategic Management Journal*, **25**, 909–28.
Friedrich, R. (1982), 'In defense of multiplicative terms in multiple regression equations', *American Journal of Political Science*, **26**, 797–833.
Gilfillan, S. (1935), *Inventing the Ship*, Chicago: Follett.
Griliches, Zvi (1957), 'Hybrid corn: an exploration in the economics of technological change', *Econometrica*, **25**, 501–22.
Gulati, Ranjay (1995), 'Social structure and alliance formation patterns: a longitudinal analysis', *Administrative Science Quarterly*, **40**, 619–52.
Hägerstrand, Torsten (1953 [1967]), *Innovation Diffusion as a Spatial Process*, Chicago: University of Chicago Press.
Hall, Bronwyn H., Adam B. Jaffe and Manuel Trajtenberg (2001), 'The NBER patent citations data file: lessons, insights and methodological tools', National Bureau of Economic Research Working Paper No. 8498.
Hansen, Morten T. (1999), 'The search-transfer problem: the role of weak ties in sharing knowledge across organization subunits', *Administrative Science Quarterly*, **44**, 82–111.
Hargadon, Andrew (1998), 'Diffusion of innovations', in Richard C. Dorf (ed.), *The Technology Management Handbook*, Boca Raton, FL: CRC/IEEE.
Hedström, Peter (1994), 'Contagious collectives: on the spatial diffusion of Swedish trade unions', *American Journal of Sociology*, **99**, 1157–79.

Henderson, Rebecca and Iain Cockburn (1996), 'Measuring competence? Exploring firm effects in pharmaceutical research', *Strategic Management Journal*, **15**, 63–84.
Homans, George C. (1950), *The Human Group*, New York: Harcourt, World and Brace.
Irwin, Douglas A. and Peter J. Klenow (1994), 'Learning-by-doing spillovers in the semiconductor industry', *Journal of Political Economy*, **102**, 1200–27.
Jaffe, Adam B., Manuel Trajtenberg and Rebecca Henderson (1993), 'Geographic localization of knowledge spillovers as evidenced by patent citations', *Quarterly Journal of Economics*, **108**, 577–98.
Jordan, Kathleen and Michael Lynch (1992), 'The sociology of a genetic engineering technique: ritual and rationality in the performance of the "plasma prep"', in Adele Clarke and Joan Fujimara (eds), *The Right Tools for the Job: At Work in the Twentieth Century Life Sciences*, Princeton, NJ: Princeton University Press, pp. 77–114.
Kauffman, Stuart A. (1993), *The Origins of Order*, New York: Oxford University.
King, Gary and Langehe Zeng (2001), 'Logistic regression in rare events data', *Political Analysis*, **9**, 137–63.
Kogut, Bruce and Udo Zander (1992), 'Knowledge of the firm, combinative capabilities, and the replication of technology', *Organization Science*, **3**, 383–97.
Kolmogorov, Andrey N. (1965), 'Three approaches to the quantitative definition of information', *Problems of Information Transmission*, **1**, 1–17.
Krugman, Paul R. (1991), *Geography and Trade*, Cambridge, MA: MIT Press.
Lanjouw, Jean O. and Mark Schankerman (2004), 'Patent quality and research productivity: measuring innovation with multiple indicators', *Economic Journal*, **114**, 441–65.
Lazarsfeld, Paul F., Bernard Berelson and Hazel Gaudet (1944), *The People's Choice: How the Voter Makes Up His Mind in a Presidential Election*, New York: Duell, Sloan, and Pearce.
Levin, Richard C., Alvin K. Klevorick, Richard R. Nelson and Sidney G. Winter (1987), 'Appropriating the returns from industrial research and development', *Brookings Papers on Economic Activity*, **3**, 783–820.
Levinthal, Daniel (1997), 'Adaptation on rugged landscapes', *Management Science*, **43**, 934–50.
Lim, Kwanghui (2009), 'The many faces of absorptive capacity: spillovers of copper interconnect technology for semiconductor chips', *Industrial and Corporate Change*, **18**, 1249–84.
Lippman, Steve and Richard Rumelt (1982), 'Uncertain imitability: An analysis of interfirm differences in efficiency under competition', *Bell Journal of Economics*, **13**, 418–38.
McEvily, Susan K. and Bala Chakravarthy (2002), 'The persistence of knowledge-based advantage: an empirical test for product performance and technological knowledge', *Strategic Management Journal*, **23**, 285–306.
McPherson, J. Miller, Pamela A. Poplielarz and Sonja Drobnic (1992), 'Social networks and organizational dynamics', *American Sociological Review*, **57**, 153–70.
McPherson, J. Miller, Lynn Smith-Lovin and James M. Cook (2001), 'Birds of a feather: homophily in social networks', *Annual Review of Sociology*, **27**, 415–44.
Mahajan, Vijay, Eitan Muller and Frank M. Bass (1990), 'New product diffusion models in marketing: a review and directions for research', *Journal of Marketing*, **54**, 1–26.
Mansfield, Edwin (1968), *Industrial Research and Technological Innovation*, New York: W.W. Norton.
Manski, Charles F. and Steven R. Lerman (1977), 'The estimation of choice probabilities from choice based samples', *Econometrica*, **45**, 1977–88.
March, James G. and Herbert A. Simon (1958), *Organizations*, Cambridge, MA: Blackwell.
Marsden, Peter V. and Noah E. Friedkin (1993), 'Network studies of social influence', *Sociological Methods and Research*, **22**, 127–51.
Marshall, Alfred (1890), *Principles of Economics*, London: Macmillan.
Mead, C. and L. Conway (1980), *Introduction to VSLI Systems*, Reading, MA: Addison-Wesley.
Nelson, Richard R. and Sidney G. Winter (1982), *An Evolutionary Theory of Economic Change*, Cambridge, MA: Belknap Press.
O'Sullivan, Alan (2001), 'Achieving modularity: generating design rules in an aerospace design-build network', Working Paper, University of Ottawa.
Owen-Smith, J. and W.W. Powell (2004), 'Knowledge networks as channels and conduits: the effects of spillovers in the Boston biotechnology community', *Organization Science*, **15**, 5–21.
Park, Robert E. (1926), 'The urban community as a spatial pattern and a moral order', in E.W. Burgess (ed.), *The Urban Community*, Chicago: University of Chicago, pp. 3–18.
Perrow, Charles (1984), *Normal Accidents: Living with High-Risk Technologies*, New York: Basic Books.
Pisano, Gary, Amy Edmondson and Richard Bohmer (2001), 'Organizational differences in rates of learning: evidence from the adoption of minimally invasive cardiac surgery', *Management Science*, **47**, 752–68.
Podolny, Joel M. (1994), 'Market uncertainty and the social character of economic exchange', *Administrative Science Quarterly*, **39**, 458–83.
Polanyi, Michael (1966), *The Tacit Dimension*, New York: Anchor Day.
Porter, Michael E. and Jan W. Rivkin (1999), 'Matching Dell', Harvard Business School Case 799–158.

Prentice, R.L. and R. Pyke (1979), 'Logistic disease incidence models and case-control studies', *Biometrika*, **66** (3), 403–11.
Reed, Ricard and Robert J. DeFillippi (1990), 'Causal ambiguity, barriers to imitation, and sustainable competitive advantage', *Academy of Management Review*, **15**, 88–102.
Reinganum, Jennifer F. (1981), 'Market structure and the diffusion of new technology', *Bell Journal of Economics*, **12**, 618–24.
Rivkin, Jan W. (2000), 'Imitation of complex strategies', *Management Science*, **46**, 824–44.
Rivkin, Jan W. (2001), 'Reproducing knowledge: replication without imitation at moderate complexity', *Organization Science*, **12**, 274–93.
Rogers, Everett M. (1995), *Diffusion of Innovations*, 4th edn, New York: Free Press.
Romer, Paul (1987), 'Growth based on increasing returns due to specialization', *American Economic Review*, **77**, 56–62.
Ryan, Bryce and Neal C. Gross (1943), 'The diffusion of hybrid seed corn in two Iowa communities', *Rural Sociology*, **8**, 15–24.
Sampat, Bhaven N. (2004), 'Examining patent examination: an analysis of examiner and applicant generated prior art', Working Paper, Georgia Institute of Technology.
Saxenian, Annalee (1994), *Regional Advantage*, Cambridge, MA: Harvard University Press.
Scherer, Frederic M. (1984), *Innovation and Growth: Schumpeterian Perspectives*, Cambridge, MA: MIT Press.
Schumpeter, Joseph (1939), *Business Cycles*, New York: McGraw-Hill.
Scott, A.J. and C.J. Wild (1997), 'Fitting regression models to case-control data by maximum likelihood', *Biometrika*, **84** (1), 57–71.
Simon, Herbert A. (1962), 'The architecture of complexity', *Proceedings of the American Philosophical Association*, **106**, 467–82.
Singh, Jasjit (2005), 'Collaboration networks as determinants of knowledge diffusion processes', *Management Science*, **51**, 756–70.
Smith, John K. and David A. Hounshell (1985), 'Walter H. Corrothers and fundamental research at DuPont', *Science*, **229**, 436–42.
Sorenson, Olav (2004), 'Social networks, informational complexity and industrial geography', in D. Fornahl, C. Zellner and D. Audretsch (eds), *The Role of Labour Mobility and Informal Networks for Knowledge Transfer*, Berlin: Springer-Verlag, pp. 79–96.
Sorenson, Olav (2005), 'Social networks and the persistence of clusters: evidence from the computer workstation industry', in Stefano Breschi and Franco Malerba (eds), *Clusters, Networks and Innovation*, New York: Oxford University Press, pp. 297–316.
Sorenson, Olav and Pino G. Audia (2000), 'The social structure of entrepreneurial activity: geographic concentration of footwear production in the U.S., 1940–1989', *American Journal of Sociology*, **106**, 424–61.
Sorenson, Olav and Lee Fleming (2004), 'Science and the diffusion of knowledge', *Research Policy*, **33**, 1615–34.
Sorenson, Olav and Toby E. Stuart (2001), 'Syndication networks and the spatial diffusion of venture capital investments', *American Journal of Sociology*, **106**, 1546–88.
Stern, Scott (2001), Personal communication.
Strang, David and Sarah A. Soule (1998), 'Diffusion in organizations and social movements: from hybrid corn to poison pills', *Annual Review of Sociology*, **24**, 265–90.
Stuart, Toby E. and Olav Sorenson (2003), 'The geography of opportunity: spatial heterogeneity in founding rates and the performance of biotechnology firms', *Research Policy*, **32**, 229–53.
Szulanski, Gabriel (1996), 'Exploring internal stickiness: impediments to the transfer of best transfer within the firm', *Strategic Management Journal*, **17** (Winter Special Issue), 27–43.
Teece, David J. (1977), 'Technology transfer by multinational firms: the resource cost of transferring technological know-how', *Economic Journal*, **87**, 242–61.
Tomz, Michael (1999), Relogit (Stata ado file), available at http://gking.harvard.edu/stats.shtml.
Trajtenberg, Manuel (1990), 'A penny for your quotes: patent citations and the value of innovations', *Rand Journal of Economics*, **21**, 172–87.
Tushman, Michael and Philip Anderson (1986), 'Technological discontinuities and organization environments', *Administrative Science Quarterly*, **31**, 439–65.
Ulrich, Karl (1995), 'The role of product architecture in the manufacturing firm', *Research Policy*, **24**, 419–40.
Usher, Abbott (1954), *A History of Mechanical Invention*, Cambridge, MA: Dover.
von Hippel, Eric (1988), *The Sources of Innovation*, New York: Oxford University.
Weick, Karl E. (1976), 'Educational organizations as loosely coupled systems', *Administrative Science Quarterly*, **21**, 1–19.
Winter, Sidney G. (1995), 'Four Rs of profitability: rents, resources, routines, and replication', in C.

Montgomery (ed.), *Resource-Based and Evolutionary Theories of the Firm: Towards a Synthesis*, Boston: Kluwer, pp. 147–78.
Womack, James P., Daniel T. Jones and Daniel Roos (1990), *The Machine that Changed the World*, New York: Rawson.
Zander, Udo and Bruce Kogut (1995), 'Knowledge and the speed of transfer and imitation of organizational capabilities: an empirical test', *Organization Science*, **6**, 76–92.
Zimmerman, Martin B. (1982), 'Learning effects and the commercialization of new energy technologies: the case of nuclear power', *Bell Journal of Economics*, **13**, 297–310.
Zucker, Lynne G., Michael R. Darby and Marilyn B. Brewer (1997), 'Intellectual human capital and the birth of U.S. biotechnology enterprises', *The American Economic Review*, **88**, 290–306.

Appendix: Simulation of knowledge flow

A simple simulation of knowledge flow serves two purposes. It clarifies further why the value of social proximity reaches its peak in the transfer of knowledge with intermediate interdependence. It also identifies the range of empirical results consistent with our theoretical model. Specifically, the theoretical model yields a unique prediction about the impact of knowledge interdependence on the *gap* between citation rates of socially close actors and socially distant actors, but can encompass a range of findings about the effect of interdependence on close-actor citation rates alone or on distant-actor rates alone.

Model

Superstructure The model employs Kauffman's (1993) NK approach, which a growing number of researchers have used to simulate technological or organizational search. The simulation unfolds as follows. First, we choose two parameters: N, the number of components or ingredients that comprise a piece of knowledge, and K, the degree to which those components interact in determining the utility of the knowledge. Using techniques described below, a simulation then generates – in a stochastic manner – a mapping from each possible way of configuring the N components (i.e. each conceivable recipe) to a measure of utility. One can visualize the mapping as a landscape in a high-dimensional space. Each discrete component constitutes a 'horizontal' axis, and each possible manner of using the component represents a point along that axis. The vertical axis records the usefulness of the resulting piece of knowledge.

Next, we assume that some firm has happened upon the most useful possible piece of knowledge – the best way to configure the components (i.e. the template described in the main chapter).[1] Two new parties then enter the landscape. One party, a close actor, has access to the owner of the template, presumably through a social tie, while the other, a distant actor, cannot access the original template through his social network. Both strive to rediscover the original success – the model's equivalent to the efforts to receive and build on knowledge discussed in the main text. Thanks to its superior access to the template, the close actor enjoys an advantage in this search process. The close actor may begin its search closer to the original success, reflecting the better information it receives or its superior ability to interpret the transmission. Or, it may move toward the success with greater speed and accuracy, reflecting its ability to seek advice from the owner of the template. The simulation models these mechanisms and records the relative success of the close actor and the distant actor in rediscovering the original piece of knowledge.

Following this first iteration, the simulation generates a second mapping that, though it differs in its particulars, has the same degree of interdependence as the first. A second pair of close and distant actors tackle the second problem, and the program records their relative success. The simulation iterates through this process hundreds of times. From the repetition emerges a profile of how close and distant actors fare relative to one another for a given degree of interdependence. We then adjust K, the parameter that governs interdependence, and repeat the process. By doing so, we build an understanding of how interdependence affects the relative ability of close and distant actors to rediscover the original success.

This description of the model's superstructure leaves two aspects of the simulation unspecified: how we generate landscapes and how actors search to rediscover the original success.

Generation of landscapes Each piece of knowledge consists of N components, and each component j, $j \in \{1, 2, \ldots, N\}$, can be configured in two ways. Hence a particular piece of knowledge $\boldsymbol{s}$ is an N-vector $\{s_1, s_2, \ldots, s_N\}$ with $s_j \in \{0, 1\}$. In the knowledge germane to a chemical process, for instance, component j might indicate the inclusion or exclusion of a particular catalyst. Similarly, a string of four components could represent which of $2^4 = 16$ shades a heated mixture must turn before being removed from a flame. For any set of components, 2^N possible pieces of knowledge (recipes) exist. We assign a utility value to each of these as follows. Assume that each component contributes C_j to utility. C_j, depending not only on the configuration, 0 or 1, of component j, but also on the configuration of K other randomly assigned components: $C_j = C_j(s_j; s_{j1}, s_{j2}, \ldots, s_{jK})$. For each possible realization of $(s_j; s_{j1}, s_{j2}, \ldots, s_{jK})$, we draw a contribution C_j at random from a uniform distribution between 0 and 1. The overall utility associated with a piece of knowledge, then, averages across the N contributions:

$$U(\boldsymbol{s}) = [C_j(s_j; s_{j1}, s_{j2}, \ldots, s_{jK})] / N.$$

K, the parameter that governs interdependence, ranges from 0 to $N - 1$.[2] $K = 0$ corresponds to a simple situation in which the contribution of each component depends only on the configuration of that component. $K = N - 1$ captures a complex setting in which the contribution of each component depends delicately on the configuration of every other component.

Once the modeler sets N and K and the simulation generates a particular landscape (i.e. a utility $U(\boldsymbol{s})$ for each of the 2^N possible pieces of knowledge), the simulation notes the piece of knowledge $\boldsymbol{s}^*$ that produces the greatest utility, which serves as a template in subsequent search efforts.

Search A modeled close actor and a modeled distant actor enter the landscape, and each struggles to rediscover the original success. Reflecting the reasoning early in the main text, neither begins precisely atop the peak at $\boldsymbol{s}^*$. Rather, each receives an imperfect transmission of the effective knowledge and begins some distance d from $\boldsymbol{s}^*$ (i.e. d of its N components differ from $\boldsymbol{s}^*$). It must then correct its understanding through search. We consider two types of search. A party involved in incremental search adjusts one component, accepts the adjustment if it produces an improvement in utility, and ceases to search when no improvement opportunities remain. A party engaged in long-jump search changes multiple decisions at once, leaping toward $\boldsymbol{s}^*$. Its leap typically misses the target; it replicates each component of $\boldsymbol{s}^*$ with probability θ. $\theta < 1$ reflects imperfect access to the template. After its leap, the long jumper improves incrementally until it exhausts opportunities. Note that either type of search could terminate on a local peak, instead of at $\boldsymbol{s}^*$.

Though both parties have imperfect access, the close actor has better access because of his or her social proximity to the original success, which serves as a template. We model the impact of social proximity in three ways. The proximate actor may begin the search closer to $\boldsymbol{s}^*$ ($d_{close} < d_{distant}$), leap toward $\boldsymbol{s}^*$ with greater accuracy ($\theta_{close} > \theta_{distant}$), or – in leaping toward s^* – may know which components are 'right' and which are 'wrong'. These benefits reflect both the more accurate transmission the close actor receives originally and his or her ability to consult with the owner of the template while trying to correct the original transmission.

Interdependence and the landscape

Much of the intuition of the results flows from an understanding of the impact of K on the topography of the typical landscape. Four effects strike us as especially germane.[3] First, as K increases, the landscape shifts from being smooth and single-peaked to being rugged and multi-peaked. When $K = 0$, the N components contribute independently to knowledge utility. In that situation, alteration of a single component changes the contribution of that component alone. From any initial location on a landscape, then, a close or distant actor can climb to the global peak via a series of utility-improving, single-component tweaks to its knowledge. In contrast, when $K = N - 1$, every component influences the contribution of every other component. Then a small step on the landscape – a change in a single component – alters the contributions of all N components. Consequently, adjacent pieces of knowledge have altogether uncorrelated utilities, producing a very rugged surface with many local peaks.

Second, as K rises, not only do local peaks proliferate, but also the height of the average peak declines. As the web of connections across components thickens, it becomes possible to exhaust opportunities for incremental improvement even at low levels of performance. Hence, interdependence decreases the fruitfulness of incremental search.

Third, though the height of the average peak falls as K rises, the heights of the highest peaks rise with K. When components interact with one another more richly, the amount of variety attainable by mixing and matching components increases, and the quality of the best combination within that variety improves. Rugged landscapes, though challenging to navigate, offer greater fertility than smooth ones – in other words, they more likely produce at least one exceptional peak. More mechanically, recall that we drew a contribution C_j for each possible realization of $(s_j; s_{j1}, s_{j2}, \ldots, s_{jK})$. The number of possible contributions for each component (2^{K+1}) rises sharply with K, increasing the available variety.

Finally, as K increases, the high peaks on the typical landscape spread apart from one another, shifting from a situation in which peaks cluster in mountain ranges to one in which peaks spread uniformly across the terrain.[4] With greater interdependence, high peaks carry less and less information about the location of other high peaks. This effect undermines long-jump search, decreasing the likelihood that a jump that aims for but misses the global peak will nonetheless land on high ground.

Simulations and results

Percentage of template performance attained We explored the model under a wide variety of assumptions regarding d_{close}, $d_{distant}$, θ_{close}, and $\theta_{distant}$. ($N = 12$ throughout. All results average over 100–200 landscapes.) Results remained similar throughout the parameter space so we report only a handful of representative cases here (see Rivkin, 2001, for further robustness checks). Figures 15A.1–15A.3 show, as a function of K, the utility attained by the close actor and the distant actor as a percentage of the utility of the template. Figure 15A.1 considers the case of incremental search with $d_{close} = 4$ and $d_{distant} = 10$. Figure 15A.2 examines the case of long-jump search with $d_{close} = d_{distant} = 12$, $\theta_{close} = 0.6$, and $\theta_{distant} = 0.4$. Figure 15A.3 considers a situation in which both parties start with a poor replica ($d_{close} = d_{distant} = 12$), each tweaks uphill to a local peak, each then leaps toward $\boldsymbol{s^*}$ with equal accuracy ($\theta_{close} = \theta_{distant} = 0.5$), but in taking the leap, only the close actor knows which of its components matches the components of $\boldsymbol{s^*}$.

In all cases, greater interdependence undermines both close and distant actors, but the greatest gap between the two arises at an intermediate level of K. To see why, consider three situations:

- When $K = 0$, the close actor has no advantage at all. The smooth landscape allows both firms to discover the global peak eventually.
- As K rises, a gap emerges between the close actor's performance and that of the distant actor. The landscape is rugged enough that the distant actor becomes stranded far from the global peak, and peaks cluster enough that average peak height declines with distance from the global peak. The landscape is sufficiently smooth and clustered, however, that the close actor – starting near $\boldsymbol{s^*}$ or leaping toward $\boldsymbol{s^*}$ accurately – can scale $\boldsymbol{s^*}$ or a nearby, nearly-as-high peak.
- As K approaches N, the gap closes. The landscape becomes so rugged that even the close actor becomes stranded on a peak other than $\boldsymbol{s^*}$. The close actor may finish closer to $\boldsymbol{s^*}$ than the distant actor does, but with high peaks no longer clustered together, this proximity has little benefit. When components depend on each other delicately, superior but slightly imperfect access to the template has little more value than highly imperfect access.

Adjusting for frequency of attempts The results so far report the knowledge-rediscovery success of the close actor versus the distant actor *conditional on both parties attempting to rediscover the knowledge embodied in the original success*. In our empirical tests, however, we examine the rates of patent citations by close and distant actors. We interpret these rates as an indication of the number of times the knowledge underlying the focal patent has been received and built upon. Accordingly, the rates reflect not only the degree of success conditional on an attempt at rediscovery being made, *but also the frequency with which attempts are made*. If, for instance, the frequency of attempts varies systematically with K, then the graphs of close- and distant-actor patent counts versus K might reveal shapes that differ in important ways from the pattern shown in Figures 15A.1–15A.3. In this light, we consider three scenarios.

First and most simply, suppose that the number of attempts made by close and distant actors is independent of K. Then we would expect the graphs of citation rates to resemble Figures 15A.1–15A.3 without modification. In other words, the frequency of both close-actor and distant-actor citation would decline with K, and the maximal difference would occur at intermediate K.

Second, assume that the number of attempts made by socially close and distant actors increases in proportion to the utility associated with the original success (i.e. more useful pieces of knowledge attract more attempts at rediscovery). Recall that the utility of the best piece of knowledge – the height of the global peak on the landscape – rises with K, reflecting the greater variety that comes from mixing and matching more interdependent

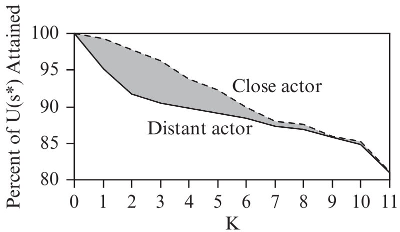


Figure 15A.1: Incremental search

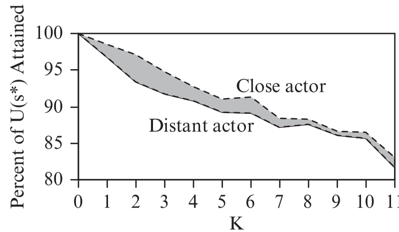


Figure 15A.2: Long-jump search

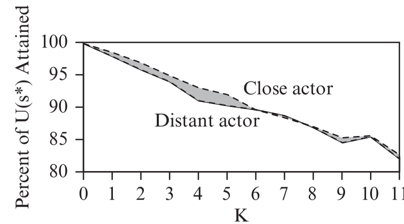


Figure 15A.3: Long jumps with vs. without knowledge of errors

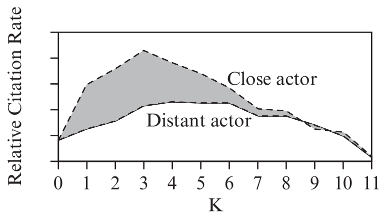


Figure 15A.4: Incremental search with number of attempts proportional to utility of template

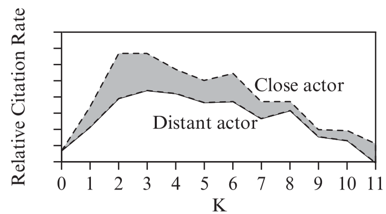


Figure 15A.5: Long-jump search with number of attempts proportional to utility of template

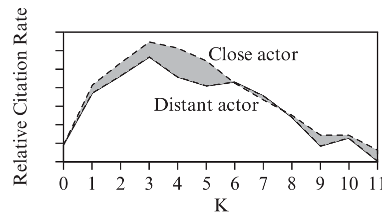


Figure 15A.6: Long jumps with vs. without knowledge, number of attempts proportional to utility of template

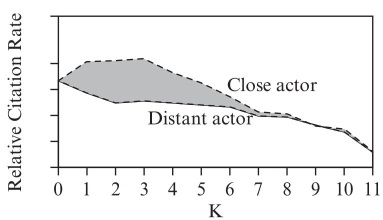


Figure 15A.7: Incremental search with number of attempts proportional to expected utility of attempt

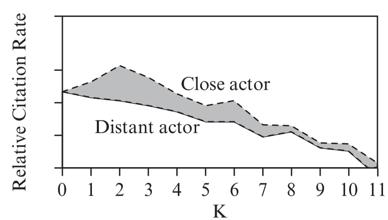


Figure 15A.8: Long-jump search with number of attempts proportional to expected utility of attempt

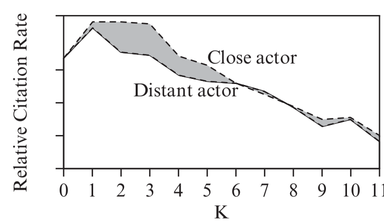


Figure 15A.9: Long jumps with vs. without knowledge, number of attempts proportional to expected utility

Note: Parameter values for each simulation are given in text. Each data point is an average over 100 landscapes.

components. When we adjust Figures 15A.1–15A.3 to incorporate more frequent rediscovery efforts on high-K landscapes, the citation pattern shifts to that shown in Figures 15A.4–15A.6. In contrast to Figures 15A.1–15A.3, the frequency of close- and distant-actor citation now rises at first, reflecting the fertility of higher-K landscapes, but then declines. In line with Figures 15A.1–15A.3, the largest gap between close-actor and distant-actor citation arises for knowledge of intermediate interdependence.

Finally, suppose that the number of attempts made by close and distant actors reflects the utility that each expects to attain in a rediscovery attempt. In deciding whether to engage in an attempt, parties not only understand that potential utility increases with K, but they also adjust for the odds that they succeed. For instance, distant actors understand they have lower odds of success and therefore make fewer attempts than do close actors. When we adjust Figures 15A.1–15A.3 in this manner, we project the citation pattern shown in Figures 15A.7–15A.9. Now the distant-actor citation rate declines monotonically with K while the close-actor citation rate has an inverted-U shape. Still, the gap between the two reaches its peak at an intermediate value of K.

In sum, the robust prediction of our theory concerns *the gap between citation rates of close and distant actors*, not close-actor citation rates by themselves or distant-actor citation rates alone. The gap between the two citation rates should have an inverted-U relationship with respect to interdependence.

Notes

1. Our focus on the global maximum simplifies the simulation, but the results remain qualitatively robust to a wide range of alternative assumptions.
2. Note that the empirically derived measure of coupling in the main text, **k**, corresponds to the parameter K in the simulation model, but the two differ at least in terms of scaling. For more on this relationship, see Note 9.
3. Kauffman (1993) explores these effects further.
4. For the intuition behind this effect, see Rivkin (2001), p. 283.

16 The geography of knowledge spillovers: the role of inventors' mobility across firms and in space

Stefano Breschi, Camilla Lenzi, Francesco Lissoni and Andrea Vezzulli

1. Introduction

In the past 20 years, research on the geography of innovation has revolved largely around the concept of 'localized knowledge spillovers' (hereafter LKSs). LKSs are 'pure externalities' (Griliches, 1992): they exist insofar as scientific and technological knowledge may escape its producer's control, and yet diffuse only locally. LKSs may explain why innovation activities are often found to be spatially clustered (Feldman, 1999).

For long supported only by circumstantial evidence, the LKS hypothesis was first tested by Jaffe, Trajtenberg and Henderson (1993; hereafter JTH). The three authors argued that knowledge spillovers may be measured by the 'citations to prior art' contained in most patent documents, and produced a statistical experiment showing that such citations come disproportionately from the same geographical area of the cited patents. The experiment requires matching each citing patent to a control one, with the same application date and technological classification, in order to compare their location in space.

The JTH experiment has become a classical reference for most empirical work on the geography of innovation, both within mainstream economics and for unorthodox approaches, such as evolutionary and institutional ones. However, its interpretation as proof of the existence of LKSs relies on the twin assumptions that scientific and technological knowledge is largely tacit, so that face-to-face contacts are the necessary vehicle for its diffusion, and that geographical proximity is a necessary condition for those contacts to take place. As a result, JTH's work treats *geographical proximity* as a proxy for *social proximity*, which inventors may derive from professional collaboration or common affiliation to companies, technical and scientific societies, or former institution of higher education. A limitation of this strategy is that it makes it difficult to distinguish between different social ties, according either to their nature (for example, professional vs. friendly) or strength (such as older vs. more recent ties).

Various attempts have been made to overcome this limitation. Agrawal et al. (2006) have tested whether inventors who move from one company to another and across different locations still pass on knowledge to former colleagues active in the cities they have left. Breschi and Lissoni (2009) and Singh (2005) have resorted to a more direct approach, which consists of measuring professional ties between inventors resulting from co-invention data, and in applying to the resulting data some standard tools of social network analysis.

In this chapter we put together the two strands of research. In particular, we explore the social network of inventors who are mobile in space, and test whether its geographical spread may help in explaining the spatial distribution of the knowledge flows generated by inventors who move both across organizations and geographical locations.

This research effort is relevant for evolutionary economic geography for at least two reasons. First, it does not merely pay lip service to the conceptualization of knowledge as tacit and situated, but it explores in depth the implications of such conceptualization. In order to do so, it avoids taking the logical shortcut of assuming that tacitness necessarily implies the localization of knowledge flows, as often happens in the applied literature based on both the institutionalist and mainstream economic approaches (Boschma and Frenken, 2006; Breschi and Lissoni, 2001).

Second, it promotes the social network as the main unit of analysis for knowledge diffusion, instead of the city or region. By doing so, it shares one of the distinctive features of evolutionary economic geography outlined by Boschma and Frenken (2006), which consists of studying whether the spatial relations between economic agents matter for technological change, and how those relations and their importance may change over time.

In section 2 we sum up the key details of the JTH experiment, and of the related experiment by Agrawal et al. (2006). In section 3 we show that patents contain enough information to measure social distance quite accurately, recall a few notions of social network analysis, and apply them to a large set of patent applications in organic chemistry, pharmaceuticals, and biotechnology, signed by US inventors. In the same section, we provide descriptive evidence on the extent of inventors' mobility across firms and space, and on the resulting shape and geographical features of the social network of inventors.

In section 4 we use our patent sample to reproduce Agrawal et al.'s version of the JTH experiment and to show that inventors' mobility across firms and social ties between inventors largely explain both the original JTH and Agrawal et al.'s results.

In the Conclusions we emphasize that our evidence casts some doubts on the common interpretation of citation-measured knowledge flows as pure externalities, or spillovers, and outline our future research plans.

2. The JTH experiment: methodology and interpretation

Methodology

The JTH experiment starts with the selection of a sample of *originating* (cited) patents. For each originating patent, all subsequent patents that cite it as prior art are then collected, after previous exclusion of company self-citations, that is, pairs of citing–originating patents assigned to the same company.[1] The address of inventors recorded in patent documents is then used to assign patents to a geographical area, in order to compare the locations of citing and originating patents.[2]

A *control* sample of patents is also built. Each citing patent is matched to a randomly drawn patent, with the same technology class and application date, but no citation link to the corresponding originating patent.

A test follows, which consists of comparing the frequency with which citing–originating patent pairs match geographically (in our experiment, at the city level) to the corresponding frequency for control–originating patent pairs. If the former turns out to be significantly greater than the latter, this should be interpreted as evidence of localization effects of spillovers *over and above* the agglomeration effects arising from other sources.[3]

The evidence reported by JTH shows indeed that citations are highly localized. Citing patents are up to two times more likely than the control patents to come from the same state, and up to six times more likely to come from the same metropolitan area.

Agrawal et al. (2006) propose a variation on the original JTH experiment that sheds some light on the nature and durability in time of social ties. In particular, they focus on the citations received by patents signed by 'movers', the latter being inventors who appear to have changed location over time (that is, they have signed at least two patents, reporting at least once a different address). To identify such patents, each citing–cited (control–cited) patent pair is 'unbundled' in order to obtain as many observations as the number of co-inventors listed on the originating patent; that is, from citing(control)–cited pairs, one obtains several so-called citing(control)–cited–inventor 'triples'. This allows identifying the relevant movers and their patents, whose 'triples' are then retained for running the co-location test. The latter is performed considering both the movers' 'current' location (the city where the mover resided at the time of the patent application) and the 'prior' ones (the city/cities where the mover used to reside at an earlier time, as witnessed by his or her earlier patent applications).

Agrawal et al. find that movers' patents are more heavily cited than expected not only by inventors resident in the movers' current locations, but also by those who reside in the prior ones. Among the latter, inventors with a past working spell in the same companies as the movers appear to be the majority. These results suggest that movers keep in touch with their former colleagues, or at least that they leave a lasting knowledge legacy behind them.

Interpretation and the role of social networks

The mobility of R&D scientists and engineers within a localized labour market and the existence of localized markets for technologies have both been reported by various authors as potential explanations of JTH results[4].

As for labour mobility, Almeida and Kogut (1999) have replicated the JTH exercise for each US state. They find evidence of localized knowledge flows only in those few regions (most notably, the Silicon Valley) where the intra-regional mobility of inventors across companies is high.

Markets for technologies may also explain the JTH results to the extent that co-location is encouraged by technology users' need to consult frequently with their suppliers. Research contracts signed by the same independent inventor with different companies may produce patents that appear to be unrelated in terms of ownership, but very close in terms of technological contents and geographical distance (Mowery and Ziedonis, 2001).

The above-mentioned studies suggest that patents linked by a citation may also be personally or socially linked. A personal tie occurs whenever the same inventor is responsible for two patents from two different companies, either because he or she moved from one to another, or because he or she is an independent inventor who sold ideas to both. A social tie exists whenever two inventors i and j working for two different companies have a common professional acquaintance, in this case a fellow inventor who has worked jointly on a patent both with i and j (he/she may be either an employee who has moved across the two companies, or an independent professional who has consulted or done research for both of them). Other social ties between two patents may involve a longer chain of acquaintances, such as when inventors *i* and *j* are connected by two other inventors *z* and *w* who first worked together, and then moved on to work separately with *i* and *j*, and so on.

More indirect social ties may also exist, such as serendipitous encounters between any two inventors who meet at workshops or through friends and other non-professional acquaintances. We fail to capture these ties with patent data, which record only formal collaboration instances.

In principle, social ties may or may not be concentrated in the geographical space: robust social links between inventors may convey tacit information even when the inventors have just a few chances to meet personally (witness many fruitful academic cooperation experiences), or well after the inventors have last met.

Agrawal et al.'s (2006) extension of the JTH experiment improves our understanding of what types of social tie really matter for knowledge diffusion. Still, it tells us little of who knows whom in the inventors' community. We know that patents are disproportionally concentrated in the hands of relatively few large companies, wherein inventors may or may not have a chance to meet and exchange knowledge, or get in touch with each other's ideas via chains of mutual acquaintances; at the same time, these companies may provide for the codification of their inventors' knowledge in order to help its circulation among all technical employees.

Both the original JTH experiment and its extension by Agrawal et al. (2006) make use of a database with limited information on inventors' social capital. None of the databases contains information on the inventors' social ties; as a consequence, both geographical distance and company affiliations end up being two summary proxies for all kinds of ties, and for physical distance as such.

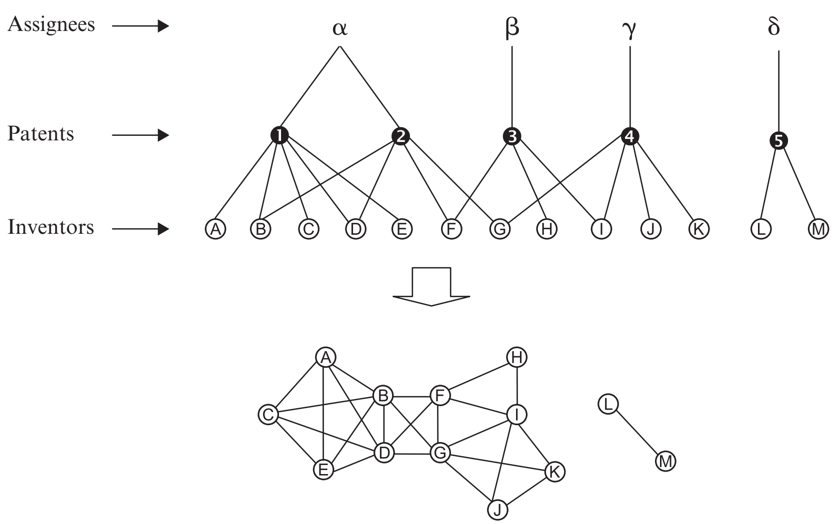
Having reclassified patents according to inventors, we will be able to show the importance of social ties as distinct from spatial localization. In particular, we will emphasize the role of a specific professional community (that of inventors) as a channel of knowledge diffusion, and show that the localization of knowledge flows is largely explained by the limited mobility in space of such a community. Thus, we will 'open the black box' of localized social ties, often evoked as the conduit of knowledge, but seldom told apart from one another and subject to measurement.

We regard our exercise as the first step towards a more comprehensive direct measurement of all social ties between inventors and, more generally, knowledge producers. It is only by detecting what types of social tie matter (in our case, we focus on collaboration ties between inventors, recorded by patent documents) that researchers will be able to explain why spatial distance is often found to limit knowledge diffusion, and possibly reach strategic or policy prescriptions.

3. Methodology and data

Social networks: definition and methodology

Our methodology exploits patent-recorded information on inventors' names, surnames, addresses, and company affiliation (Breschi and Lissoni, 2004 and 2009). The following hypothetical example illustrates the main idea (see Figure 16.1). Let's consider five patents (1 to 5) and four assignees (α, β, γ, δ). Assignee α owns two patents (1 and 2), while assignees β, γ and δ one each. Patents have been produced by 13 distinct inventors



Top: Bipartite graph of assignees (α, β, γ, δ), patents (1 to 5) and inventors (A to M), with lines linking each patent to the respective inventors and assignees.

Source: Breschi and Lissoni (2004, 2009).

Figure 16.1 Bipartite graph of patents and inventors

(A to M). For example, patent 1, assigned to company α, has been produced by a team comprising inventors A, B, C, D and E. It is reasonable to assume that, because of their collaboration in a common research project, those five inventors are *socially linked* by some kind of knowledge sharing. The existence of such a linkage can be graphically represented by drawing an undirected edge between each pair of inventors, as in the bottom part of Figure 16.1.

Repeating the same exercise for each team of inventors, we end up with a graph representing the network of all inventors. This allows us to measure how tight is the connection between each pair of patents. In order to see how, we first give a few definitions:

1. For any pair of inventors, one can measure the distance between the two by calculating the so-called *geodesic distance*. The geodesic distance is defined as the minimum number of edges that separate two distinct inventors in the network.[5] In Figure 16.1, for example, the geodesic distance between inventors A and C is equal to 1, whereas the same distance for inventors A and H is 3. While A and C shared directly their knowledge while working on patent 1, A and H are more likely to have exchanged some word-of-mouth technical information through the mediation of other actors (such as B and F).
2. Any two inventors may belong to the same *social component* or to *socially disconnected components*. A component of a graph can be defined as a subset of the entire graph, such that all nodes included in the subset are connected through some path. In Figure 16.1, for example, inventors A to K belong to the same component,

whereas inventors L and M belong to a different component. A pair of inventors belonging to two distinct components have distance equal to infinity (i.e. there is no path connecting them).
3. By *cross-firm* inventor we mean any inventor whose name has been reported in patent documents assigned to different organizations. This kind of inventor plays a fundamental role in connecting teams of inventors belonging to different organizations. For example, in Figure 16.1, inventor F worked for both company α and company β, thus connecting the team of inventors (B,D) with the team of inventors (H,I). Similarly, inventor G worked both for company α and γ, thus connecting the team (B,D,F) with the team (I,J,K).

Using these definitions, we may now turn to illustrate how the existence of a linkage between patents can be ascertained. Three possible relations exist between any pair of patents from different firms:

1. The two patents exhibit *no social connection*, such as when the inventors behind them belong to socially disconnected components.[6]
2. The two patents are linked by a *social connection*, such as when their inventors belong to the same social component. We can calculate the social distance *between patents* as the *minimum* geodesic distance between the two closest individuals from the two teams of inventors (geodesic distance).[7] As such, the social distance between two socially connected patents may vary from 1 to any positive discrete value.
3. The two patents are linked by a *personal connection*, such as when at least one inventor belongs to both patents' teams. The social distance between two personally connected patents is zero.[8]

A limitation of this approach relates to the absence of rules to establish the *decay* of social links. In fact, we know for sure when two inventors come into contact, namely when they work together on the same patent for the first time. But we cannot be sure they keep in touch (and exchange information) after that common experience, unless we find them working on more joint patents in the following years. Some contacts established through co-inventorship may be dropped by one or both parts, but we do not know which ones.

We have addressed this problem by setting a five-year maximum life span for all social ties: the tie binding two inventors who worked on the same patent at time t is cancelled at time $t + 5$, and the network is re-calculated accordingly, unless the same inventors are found to work again together at a time between t and $t + 5$. This means that for every year we calculate a different social network.[9]

Data

To implement the methodology just described, we rely on a biographical dataset of 63188 US inventors and their 66349 patent applications at the European Patent Office (EPO), filed between 1978 and 2002, in the following fields: Organic Chemistry, Pharmaceuticals, and Biotechnology (Table 16.1).[10]

The choice of the three technical fields is explained by the high degree of inter-relation among them. In a previous paper we show how it is often the case that patents whose main technological class falls in one field have secondary classes in one or both of the

Table 16.1 US inventors' patenting activity at EPO, 1978–2002 (selected IPC classes)

	1978–1985	1986–1995	1996–2002	*1978–2002*
Patents				
Pharmaceuticals (1)	1784	8739	9360	*19883*
Organic chemistry (2)	6152	13433	9845	*29430*
Biotechnologies (3)	1525	7013	8498	*17036*
Total	*9461*	*29185*	*27703*	*66349*
Inventors				
Pharmaceuticals (1)	2409	11897	14494	*24762°*
Organic chemistry (2)	5679	15289	15640	*29799°*
Biotechnologies (3)	2145	10549	12787	*22066°*
Total	*9430**	*32215**	*36661**	*63188*°*

Notes:
(1) IPC: A61K = preparations for medical/dental/toilet purposes
(2) IPC: C07, excl.C07B
(3) IPC: C12M-S = biochemistry; microbiology; enzymology; genetic engineering
* Total no. of inventors < sum of tech. classes, as one inventor may patent across classes.
° Total no. of inventors < sum of time intervals, as one inventor may patent at different times.

others (Breschi et al., 2003). Here we emphasize that many inventors active in one field also sign patents in one or both of the other fields. This explains why the total number of inventors in our sample (last line of Table 16.1) is smaller than the sum of all inventors in each technological class, whatever time period we observe. In particular, in between 1978 and 2002, 13439 inventors (21 per cent of the total sample) have patented in more than one of the three fields.

The selected inventors represent the universe of all US inventors listed on EPO documents in the selected fields and years. The resulting affiliation network of patents, applicants and inventors, is therefore comprehensive of all social ties established through joint inventive activity, as it is the one-mode projection of the same network onto just inventors. From this network we have derived measures of social distance between all patents in the sample.

A major problem with measuring the geographic dispersion of patents and patent citations relates to the way patents are assigned to locations. Patent documents report the postal address of each inventor, from which we derive the MSA (metropolitan statistical area) and state location of each inventor in each year of activity.

However, patents can have multiple inventors, each one with a different address. Therefore, the location of patents in geographic space cannot be resolved in an unequivocal way. In case of multiple inventors, JTH assigned each patent to the country/state in which pluralities of inventors resided, with ties assigned arbitrarily. Here, we take a slightly different approach and argue that two patents match geographically to the extent that they share *at least* one inventor's location.

Descriptive statistics

The vast majority (73 per cent) of all inventors in organic chemistry, pharmaceuticals, and biotechnology are found to sign patents for only one company throughout their

*Table 16.2 Inventors' activity across assignees, MSAs**

No. of assignees joined	No. of inventors (% of all inv.)	% of inventors active in >1 MSA
1	46458 (73.52)	2.3
>1	16730 (26.48)	28.4
of which		
2	*10645 (16.85)*	22.8
3	*3679 (5.82)*	34.1
4	*1439 (2.28)*	40.3
5	*562 (0.89)*	46.4
6–10	*392 (0.62)*	46.7
>10	*13 (0.02)*	53.8
All inventors	63188 (100)	9.2

Note: * The table reports the distribution of all inventors in the sample, according to the number of different assignees for which they have recorded patents over the period examined. The calculation includes patents registered by individual inventors, i.e. not assigned to organizations. The calculation includes also patents that have been co-assigned to different organizations.

inventive career (which, in many cases, is limited to only one patent). Most 'cross-firm' inventors sign patents for just two different assignees (17 per cent of all inventors; 64 per cent of cross-firm inventors); only a very few of them sign patents for more than five assignees (0.6 per cent of all inventors; 4 per cent of cross-firm inventors; Table 16.2, second column).

If inventors' cross-firm activity looks limited, mobility in space looks even more limited. Only 28.4 per cent of all cross-firm inventors (9.2 per cent of all inventors) have been active in more than one MSA. Figures grow with the number of assignees the inventors have signed patents for, and go over 40 per cent for cross-firm inventors with four assignees or more. The latter, however, are very few. Almost no inventor has been active in more than two MSAs.[11]

Table 16.3 reports some descriptive statistics for the one-mode network of inventors. By construction, the size of the network changes over time (see the social networks subsection above), with a tendency to expand because of the 'patent explosion' of the 1990s (Hall, 2004). Table 16.3 shows how the network of inventors exhibits a characteristic typical of 'small world' network of scientists, namely the asymmetry in size between the principal component (which in 1999 includes 46 per cent of the inventors) and all the other components (in 1999, the second largest component collects only 0.9 per cent of the inventors).[12]

Inventors from different companies are linked one to another by cross-firm inventors. To the extent that the latter do not move much across cities the resulting social network will be concentrated in space. Table 16.4 confirms this intuition for the principal component: C5 and C4 concentration indexes always result above 50 per cent, when calculated for active inventors (namely, those inventors in the network who sign patents in the current year). These figures are relatively stable over time, despite an increase in the number of MSAs reached by the expanding network, which suggests that peripheral areas host just a very few inventors.

*Table 16.3 Network of inventors: size of selected components; selected years**

	No. of inventors	% of inventors (excl. isolates)
1991 (no. of components = 2525)		
Principal component	3664	23.39
2nd largest "	287	1.83
Smallest "	2	0.01
1995 (no. of components = 2969)		
Principal component	7038	33.29
2nd largest "	250	1.18
Smallest "	2	0.01
1999 (no. of components = 3595)		
Principal component	14077	45.97
2nd largest "	277	0.90
Smallest "	2	0.01

Notes: * Each year's network is calculated by taking into account all patent applications for the selected technological classes over the previous five years.

Table 16.4 Concentration of inventors, by MSA; principal network component, selected years

Year	No. of inventors[a]	No. of active inventors[b]	No. of MSAs with active inventors	% of active investors in top 5 MSAs[c]
1991	3664	1509	59	59.2
1995	7038	2934	80	52.0
1999	14077	5656	103	53.8

Notes:
[a] No. of inventors in the principal component of the inventors' network. Each year's network is calculated by taking into account all patent applications for the selected technological classes over the previous five years.
[b] No. of inventors in the network having signed at least one patent application in the given year, for the selected technological classes.
[c] Top five MSAs: San Francisco, New York, Baltimore, Philadelphia, Boston.

4. The role of 'movers', and their social ties

Sampling

Following as closely as possible the JTH methodology, as adapted by Agrawal et al. (2006), we selected for this study three cohorts of *originating* patents, consisting respectively of the 1991, 1992 and 1993 patent applications that received at least one subsequent citation by the end of 1999.[13]

For each cohort of originating patents, we eliminated all applications that either received citations only from foreign organizations, or whose applicant was a US organization, *but did not* report any US inventor.[14] The choice of excluding citations from foreign companies implies that our study does not investigate the extent of international localization of patent citations. This choice was mainly dictated by data constraints, as the inclusion of citations coming from foreign organizations would have implied the construction of a worldwide network of inventors. We also removed all observations in which citing and originating patents were assigned to the same organization (i.e. company self-citations).

We additionally excluded those patent pairs (citing–cited) in which the same individual was listed in both documents (personal self-citations). Given that our focus was especially on movers, we followed Agrawal et al. (2006) and excluded these cases, to the extent that they could not represent any knowledge flow between two distinct individuals. This left us with a total of 1933 originating (i.e. cited) patents, 2868 citing patents and 3517 citing–originating patent pairs.

In order to create a sample of control patents we randomly extracted, for each citing patent, another patent with the following characteristics: no citation to the originating patent and same IPC four-digit class and priority year of the citing patent. This procedure yielded a sample consisting of 3283 control patents for a total of 3517 control–originating patent pairs (the same control patent can control for more than one citing patent).

Next, we 'unbundled' individual inventors of the originating patents: from each patent we extracted as many observations as the number of inventors listed on it, in order to obtain two 'patent triples' (citing–cited–inventor and control–cited–inventor) for each inventor. The resulting sample contains 10988 observations.

We then extracted information on the spatial and social distance between inventors within each triple in the sample.

In the first place, we asked whether at least one inventor from the citing (control) patent was located in the same current MSA of the inventor of the originating patent, and created a dummy variable accordingly. We also asked the same question regarding the prior location of the inventor, if the latter was a mover. Thus, for movers, we ended up with two dummy variables for each triple, which took a value of one in the case of geographical matching of the citing (control) patent with the mover's current and prior location, respectively.

In the second place, for each patent triple, we measured the social distance between the mover and the inventors of the citing (control) patents. Using information derived from the network of inventors, we classified all patent triple into two (mutually exclusive) groups according to the linkages connecting the individual inventors that have produced them:

- *Connected*: These were patent triples in which there was at least one inventor from the citing (respectively, control) patent that was connected to the inventor of the originating patent through a finite path in the co-invention network; we also calculated the social distance within each triple as the geodesic distance between the mover and the closest among the co-inventors of the citing (control) patent.[15]
- *Unconnected*: These are patent triples whose respective teams of inventors were not connected to each other in the co-invention network (although we could not exclude the existence of other types of informal social ties).

Table 16.5 Sample of cited, citing and control patents: summary statistics

	No. of patents	No. of patent triples	No. of inventors	No. of inventors per patent	% of connected triples
Cited	1933	–	4375	2.96	–
Citing	2868	10988	7436	3.45	35.0
Control	3283	10988	8690	3.24	24.2

Table 16.6 Socially connected citing (control) patents: frequency of geodesic distances from the originating patent

Geodesic distance	Citing	Control
1–5	24.40	15.44
6–10	49.6	45.29
11–20	23.3	36.14
>20	2.7	3.13
Mean	8.72	10.01
Median	8	9
Std deviation	4.94	4.93

Obs = 3894 citing (2656 controls).

Social ties in citing vs. control patents

Social connections between inventors are conducive to knowledge diffusion. It has been shown that the probability of observing a citation link between any pair of patents with different dates is a positive function of the existence of a personal or social tie between the two patents (Singh, 2005). In particular, social links matter when they are very close, that is when they exhibit low (≤5) geodesic distance (Breschi and Lissoni, 2004).

As a consequence, we expect that the probability of observing a personal or social link within a cited–citing patent pair is higher than the probability of observing a similar link in a cited–control pair. Table 16.5 shows that 35 per cent of citing patents are linked to originating ones through a social chain of finite length; the same value for control patents is only 24 per cent.

Table 16.6 shows that the mean geodesic distance for connected patent pairs is significantly lower for citing patents (8.72) than for control ones (10.01). While over 24 per cent of social ties between citing and originating patents is below six degrees of separation, the same figure for the control-cited pairs is under 16 per cent.

Finally, we have calculated the geographical co-location of both the citing and the control patent samples, with respect to the originating patents. As expected, closely connected patent pairs, both in the citing and the control sample, are highly co-localized with the cited one. For inventors at less than six degrees of separation, the percentage of co-location is around 29 per cent both between citing–cited patents and for control–cited patents (Table 16.7). The only major difference emerges for connected patents at more

Table 16.7 Geographic matching: % frequency for socially connected patents at the MSA level (citing and controls)

Geodesic distance	Citing (no. of triples)	Control (no. of triples)
1–5	29.4	29.5
6–10	13.5	9.6
11–20	7.0	7.0
>20	28.8	4.8
Unconnected	9.7	8.2
All	12.0	9.0

*Table 16.8 Geographical matching at MSA level: % frequency and test of proportions**

	No. of obs	Citing: % match	Control: % match	z-statistic ($P > z$)
All inventors (JTH experiment)	10988	12.01	9.05	7.16***
Only connected citing–cited patents, and related controls	4779	14.69	10.40	6.33***
Only unconnected citing–cited patents, and related controls	6209	9.95	8.00	3.80***

Notes:
* One-tail z-test on difference between co-location percentages: 2.81 (0.0234)***
Statistically significant at 1%***, 5%**, 10%*. One tail z-test $P(Z > z)$.

than 20 degrees of separation. Overall, these results suggest that, once we control for social ties, there is no reason to expect an association between geographical proximity and citations.

Analysis: the JTH experiment

As a first step in our analysis, in the 'all inventors' row of Table 16.8 we replicate the JTH exercise on our data. The second and third columns of the table report the percentage of, respectively, citing and control patents that are co-located with the originating ones, at the MSA level, while the fourth column reports the z-statistic for the difference between the two and (in brackets) the result of a 1-tail test on the hypothesis that $p_c > p_{nc}$ (where p_c is the co-location probability of citing patents, and p_{nc} the co-location probability of controls).[16]

The geographic matching rates we found for both the citing and the control patents were almost double those reported by JTH (at the city level). To the extent that localization effects are expected to fade with time, these differences may be explained by the shorter period of observation we used (9 years maximum, from 1991 to 1999, in our sample, against a maximum of 14 years in the original JTH exercise).

As for differences between citing and control patents, however, our results are very close to the original JTH ones. The proportion of citing patents co-located with originating ones is significantly greater than the proportion of control patents.

The descriptive statistics we presented in section 3 suggest that the different composition of the citing and control sample may be responsible for this result. We know that the citing patents are much more likely than the control ones to be socially connected to the originating ones, and that the cross-firm inventors responsible for the connection are most often immobile in space. Hence we expect that controlling for personal connections may reduce the observed differences in co-location rates between citing and control patents.

As the citing patent sample contains a higher number of socially connected patents with geodesic distance equal to or lower than 5 (which we know to be highly co-located with originating ones), we also expect to observe a similar result when controlling for social connections.

The 'only connected' line of Table 16.8 shows that by excluding all the unconnected patent triples from both the citing and the control samples,[17] the co-location percentages of both the citing and control samples increase. On the contrary, when considering only the unconnected patents (third line) the co-location percentages go down, although the difference between citing and control patents remains significant. When considering connected patents, the difference in co-location percentages between citing and control samples is significantly higher than when considering unconnected ones. This result suggests that professional ties, as measured in our network of inventors, may explain a sizeable part of the JTH findings. However, the persistence of some locational differences between citing and control triples leaves open the possibility that social ties other than those captured by our network may also affect inventors' knowledge exchanges, and result in patent citations.

Analysis: the role of movers

In order to focus on movers, we restricted our sample to those triples wherein the inventor of the originating patent had signed at least one patent before the originating one(s), and he/she had done so when residing in a different MSA.[18] Since movers are a very small subset of all inventors (see section 3) this restriction caused a dramatic drop in our sample size, from over 10,000 observations to just 594, resulting from 287 originating patents, 279 inventors, 477 citing patens and 499 controls. This figure represents approximately 5.5 per cent of the initial sample, a percentage that is consistent with the one found by Agrawal et al. (2006) that is approximately 6 per cent.

The distribution of patent triples between connected and unconnected ones does not change with regard to the original sample (36 per cent connected triples among citing ones; 26 per cent connected among controls). The path length distribution of connected patents, however, changes slightly, with more citing triples connected with paths shorter than six degrees of separation than in the general sample (34 per cent vs. 24 per cent; distribution for control sample does not change much; see Table 16.9).

The analysis of co-location patterns reveals some interesting results. Citing patents do not appear to be more likely to come from the same current locations of movers, no matter whether we consider the connected or the unconnected ones, or all (Table 16.10).

On the contrary, when we consider movers' prior location, we find that citing patents still tend to be more likely to be signed by inventors from those locations (11 per cent vs. 8 per cent; see Table 16.10); however, when considering social ties, this result holds (indeed is reinforced) only when patents are socially connected (14 per cent vs. 8.5 per

Table 16.9 Socially connected citing (control) patents: frequency of geodesic distances from the originating patent (movers)

Geodesic distance	Citing	Control
1–5	34.42	16.13
6–10	36.28	38.71
11–20	27.44	43.87
>20	1.86	1.29
Mean	8	10.19
Median	7	10
Std deviation	5.18	4.9

Obs = 215 citing (155 controls).

*Table 16.10 Geographical matching at MSA level, for movers' current and prior location: % frequency and test of proportions**

	No. of obs	Citing: % match	Control: % match	*z*-statistic ($P > z$)
Movers' current location				
All	594	7.74	8.42	−0.43
Connected	272	10.66	8.82	0.72
Disconnected	322	5.28	8.07	−1.42
Movers' prior location				
All	594	11.28	7.58	2.18***
Connected	272	13.97	8.46	2.04**
Disconnected	322	9.01	6.83	1.02

Notes: * Statistically significant at 1%***, 5%**, 10%*. One tail *z*-test $P(Z > z)$.

cent), and not when there is no social connection (9 per cent vs. 7 per cent, but the *z*-test suggests the difference is not significant).

These results look sharper than those achieved by Agrawal et al. (2006), both for purely statistical reasons (the smaller sample) and for substantial ones.

Our smaller sample may explain why we see the citing patents' 'co-location' premium for movers' current locations to become insignificant, while Agrawal et al. found it to diminish, but still resist.

As for prior locations, we do not only find confirmation of Agrawal et al.'s result of higher co-location of citing patents; we also find that this holds only for connected patents. This suggests that the social ties that bind movers to their prior locations originate from within their professional networks, that is they can be traced back to their former co-inventors, and to those co-inventors' collaborators up to the sixth degrees of separation or so.

5. Conclusions

Our analysis has shown that the results obtained by both the original JTH experiment and its variation by Agrawal et al. (2006) can be largely explained by the importance of social ties from within the network of inventors.

Inventors who work for different companies are responsible for a large number of citations, but are scarcely mobile in space: they move or diffuse their knowledge across different firms, but not so much across different localities. They also contribute to create social networks, which also spread knowledge across firms, but not in space. These results are confirmed by another exercise, based on the same methodology, on a sample of Italian data (Breschi and Lissoni, 2009).

Those few inventors who move not only across companies, but also across locations (the 'movers') maintain their ties with former co-inventors and their networks who are still located in the cities they have left. It is those cities, not those where the movers reside, that seem to enjoy an advantage over others in terms of access to the movers' knowledge.

Our results also raise a few substantive issues that deserve to be further discussed and investigated.

In the first place, our results qualify the original intuition of those economists and sociologists that first emphasized the tacit content of technological knowledge: knowledge always travels along with people who master it. If those people move away from where they originally learnt, researched, and delivered their inventions, knowledge will diffuse in space. Otherwise, access to it will remain constrained in bounded locations. That is, knowledge flows are localized to the extent that cross-firm activity and the resulting social networks also are localized. Why US cross-firm inventors exhibit the observed (quite limited) mobility patterns is of course an important question, but one that goes beyond the scope of the present study.

In addition, our results suggest that social ties derived from co-invention activity do not disappear when inventors move across space; on the contrary they seem to convey more knowledge than those built by movers in their new locations.

Networks of inventors, of course, capture only a tiny subset of all the relevant social contacts enabling an individual to achieve an invention. However, most inventors listed on EPO patents are professionals, and their population is much more than a tiny and unchecked sample of all the individuals who can influence them; rather, it is the most immediate and influential social environment from which inventors draw ideas and information, at least for the technical contents of their patents.

The fact that by controlling for the role of the network of inventors we manage to reduce the apparent role of spatial proximity, but not to eliminate it, suggests that other social networks, different from the professional one we considered here, may matter. In the near future, we will move in this direction by considering both patents and scientific publications, so that our network includes both inventors and pure academic scientists.

The social network approach to knowledge diffusion proposed in this chapter may be further extended to comparative analysis. Ideally, one could compare the extent of knowledge localization in different regions or nations, and explain it with the different degree of mobility and resulting network dispersion of inventors, scientists, and knowledge producers in general.

Notes

1. The exclusion of company self-citations is motivated by the fact that these citations do not represent spillovers. Originating patents with no citations are similarly excluded from the analysis, because they are supposed not to have generated any knowledge flow.

2. The rules followed by JTH to locate patents in space are indeed too complex to be summed up here. Two full paragraphs of their article are devoted to explain them (p. 585). Our rules will be slightly different and simpler, as explained below.
3. '[We estimate] the probability of a patent matching the originating patent by geographic area, *conditional* on its citing the originating patent, with the probability of a match *not conditioned on the existence of a citation link*. This noncitation-conditioned probability gives a baseline or reference value against which to compare the proportions of citations that match' (JTH, 1993, p. 581).
4. None of these explanations is entirely compatible with the 'spillover' interpretation of such results. We come back to this point in the Conclusions.
5. For technical terms from social network analysis, see Wasserman and Faust (1994).
6. With reference to Figure 16.1, this is the case, for example, of patent 5 and patent 1.
7. When two patents are socially connected, all of their inventors belong necessarily to the same social component, but not all them are at the same geodesic distance. For example, in Figure 16.1 patents 4 and 1 are socially connected, but inventor K (from patent 4) and inventor A (from patent 1) exhibit a geodesic distance of 3, while inventors G (patent 4) and B (patent 1) have a geodesic distance of just 1. G and B are the closest inventors, and it is the geodesic distance between them that we pick up as the social distance between patents 4 and 1. In other words, the social distance between the two patents is the minimum geodesic distance between their inventors. See Breschi and Lissoni (2004) for further details, and a discussion of this choice.
8. For example, inventors G, K and J (from patent 4) and H and F (from patent 3) belong to the same social component; in addition, inventor I appears both in patent 4 and 3. In the absence of I, patents 4 and 3 would be socially connect at distance 1 (the geodesic distance between G and F). The presence of I reduces this distance; we capture this reduction by setting the distance to zero.
9. We have experimented with different time windows, and the resulting networks turn out to be quite similar to those based on the five-year time frame.
10. Our data are a subset of the KEINS database, a much larger database that classifies all EPO patent applications from 1978 (EPO's first year of activity) to the current year, both by company and inventor. Fields are defined as 3- or 4-digit IPC (International Patent Classification) classes, or groups of classes. Pharmaceuticals correspond to class A61K, Organic Chemistry to class C07 (with the exclusion of C07B), and Biotechnology to classes from C12M to C12S. Companies in the KEINS database are identified by name and address. Group subsidiaries have been identified with the help of different editions of *Who Owns Whom*. As for inventors, these are identified first by assigning a unique code to all inventors with the same name, surname, and address; and then by running Massacrator©, a programme that assigns scores to any pair of inventors with the same name+surname but different address, on the basis of information suggesting the two inventors may be the same person (such as the technological class of their patents, the identity of their patents' applicants, their location in space, and the identity of their co-inventors). For details, see Lissoni et al. (2006).
11. Data available on request
12. On networks of scientists and their 'small world' properties see Newman (2001). Differently from those networks, however, ours does not exhibit low average geodesic distance: two randomly chosen inventors from the principal component are separated by more than six degrees.
13. The priority year has been used to date patent applications.
14. The nationality of inventors has been derived by the address reported in patent documents.
15. This implies that the geodesic distance assigned to each triple citing–cited–inventor or control–cited–inventor is the same for all the inventors in the originating patent team while, in principle, this distance could vary across triples (i.e. across inventors in the originating patent team). Nevertheless, the geodesic distance between each inventor of the originating patent and the relative citing (respectively, control) patent is equal either to the minimum geodesic distance between the citing–cited pair (respectively, the control–cited pair) if this inventor allows for the shortest path between the two or the minimum geodesic distance between the citing-cited pair (respectively, the control–cited pair) plus 1 (i.e. the geodesic distance of the closest inventor plus the geodesic distance of an inventor in the patent team to the closest inventor, that is by definition 1). Thus, there is a potential risk, though limited, of underestimating the geodesic distance between an inventor of the originating patent and the citing (respectively, control) patent.
16. We define: $z = (\hat{p}_c - \hat{p}_{nc})/\sqrt{\hat{p}(1-\hat{p})(1/n_c + 1/n_{nc})}$, where: $\hat{p}_c$ and $\hat{p}_{nc}$ are the sample proportion estimates for the citing and the control patents; n_c and n_{nc} are the size of the citing and control samples (in our case: $n_c = n_{nc}$); and $\hat{p} = (colocated_c + colocated_{nc})/(n_c + n_{nc})$, where $colocated_c$ and $colocated_{nc}$ are the number of co-located citing patents and controls, respectively. The 1-tail test of our interest calculates the probability attached to values higher than z from a standard normal distribution. JTH's test of proportion is slightly different, as it is based on a *t*-distributed statistic (JTH, p. 589):

$t = (\hat{p}_c - \hat{p}_{nc})/\sqrt{[\hat{p}_c(1 - \hat{p}_c) + \hat{p}_{nc}(1 - \hat{p}_{nc})]/n}$, where $\hat{p}_c$ and $\hat{p}_{nc}$ have the same meaning as above, and n is the size of both the citing and the cited samples ($n = n_c = n_{nc}$).

17. More precisely, we exclude all the citing patents that turn out to be personally connected, and the related controls; we also exclude all the control patents that turn out to be personally connected, and the citing patents they were meant to match.
18. In particular, we looked at the inventors of the originating patents and tracked backward their patent history. For each originating patent, we identified the first previous patent filed (or the set of previous patents filed in case of multiple patents in the same year), if any. We next compared the inventor's location (MSA) at the time of the originating patent and that of the first previous patent filed. If these locations differed, then we considered the inventor as a mover. In case of multiple previous patents, we relaxed this criterion and considered an inventor as a mover if at least one of the previous locations differed from the one at the time of the originating patent.

References

Agrawal, A., I. Cockburn and J. McHale (2006), 'Gone but not forgotten: knowledge flows, labour mobility, and enduring social relationships', *Journal of Economic Geography*, **6** (5), 571–91.

Almeida, P. and B. Kogut (1999), 'Localisation of knowledge and the mobility of engineers in regional networks', *Management Science*, **45** (7), 905–17.

Boschma, R.A. and K. Frenken (2006), 'Why is economic geography not an evolutionary science? Towards an evolutionary economic geography', *Journal of Economic Geography*, **6** (3), 273–302.

Breschi, S. and F. Lissoni (2001), 'Knowledge spillovers and local innovation systems: a critical survey', *Industrial and Corporate Change*, **10** (4), 975–1005.

Breschi, S. and F. Lissoni (2004), 'Knowledge networks from patent data: methodological issues and research targets', in W. Glänzel, H. Moed and U. Schmoch (eds), *Handbook of Quantitative S&T Research*, Dordrecht: Kluwer Academic Publishers.

Breschi, S. and F. Lissoni (2009), 'Mobility of skilled workers and co-invention networks: an anatomy of localised knowledge flows', *Journal of Economic Geography*, **9** (4), 439–68.

Breschi, S., F. Lissoni and F. Malerba (2003), 'Knowledge relatedness in firm technological diversification', *Research Policy*, **32** (1), 69–87.

Feldman, M.P. (1999), 'The new economics of innovation, spillovers and agglomeration: a review of empirical studies', *Economics of Innovation and New Technology*, **8**, 5–25.

Griliches Z. (1992), 'The search for R&D spillovers', *Scandinavian Journal of Economics*, **94**, Supplement, 29–47.

Hall, B. (2004), 'Exploring the patent explosion', *Journal of Technology Transfer*, **30**, 35–48.

Jaffe, A.B., M. Trajtenberg and R. Henderson (1993), 'Geographic localisation of knowledge spillovers as evidenced by patent citations', *Quarterly Journal of Economics*, **10**, 577–98.

Lissoni, F., B. Sanditov and G. Tarasconi (2006), 'The Keins database on academic inventors: methodology and contents', CESPRI Working Paper No. 181.

Mowery, D.C. and A.A. Ziedonis (2001), 'The geographic reach of market and non-market channels of technology transfer: comparing citations and licenses of university patents', NBER Working Papers No. 8568.

Newman, M.E.J. (2001), 'The structure of scientific collaboration networks', *Proceedings of the National Academy of Science USA*, **98**, 404–9.

Singh, J.H. (2005), 'Collaborative networks as determinants of knowledge diffusion patterns', *Management Science*, **51** (5), 756–70.

Wasserman, S. and C. Faust (1994), *Social Network Analysis: Methods and Applications*, Cambridge: Cambridge University Press.

17 Growth, development and structural change of innovator networks: the case of Jena

Uwe Cantner and Holger Graf

1. Introduction

The notion of collective invention has been introduced in the literature by Allen (1983) who provides evidence from the nineteenth-century iron and steel industry, where innovative success was the result of the cooperative activities of several, different actors. Anecdotal evidence such as Allen's or the various studies on Silicon Valley (e.g. Saxenian, 1994) or other success stories of regional innovation (e.g. Braczyk et al., 1998; Cooke and Morgan, 1994, Keeble et al., 1999) was enriched by studies on the regional dimension of knowledge flows (e.g. Jaffe et al., 1993). Insights into the process of innovation at the firm level (e.g. Kline and Rosenberg, 1986) or the national level (e.g. Lundvall, 1992; Nelson, 1993) strengthened the view that the functioning of innovation systems is the basis for innovation-based economic success.

Inspired by the seminal volumes by Lundvall (1992), Nelson (1993), and Braczyk et al. (1998), a large number of investigations build on the systems view in relating innovative activities and success.[1] As the systemic view of innovation is inherently dynamic and deeply grounded in evolutionary theorizing, empirical studies that take the systems view seriously should have two ingredients: heterogeneous, interacting actors and the dynamics of these interactions. Unfortunately, many of the empirical studies usually fail to account for both of these ingredients, because of the unavailability of appropriate data. In studies based on aggregate data it is easier to observe the dynamics of the system as many variables are available as time series. Many studies that account for the interactive structure and the heterogeneity of innovative actors are based on interviews or surveys that are only available at one point in time. This is not to say that system dynamics are not considered important by these authors, but most of the arguments are based on theoretical reasoning rather than on empirical observations.

As a response to this unsatisfying situation, a number of researchers in economics and economic geography started to employ social network analysis to study the interactive structures in the innovation process. The network approach offers a methodology that accounts for heterogeneity and system dynamics. However, its empirical application is subject to rather strong data requirements. Data need to be more or less complete in terms of actors and relations and for dynamic studies they have to be available over longer time periods. Information derived from patents or publications has these properties and in the application of social network analysis it provides a graphical representation of the actors and their relations in the system under investigation. For example, networks of co-authorship are used to analyse the development of scientific communities (e.g. Barabasi et al., 2002, Moody, 2004), networks of co-invention can help us understand the evolution of local clusters (e.g. Fleming et al., 2006), and citation networks provide information about the flow of knowledge (e.g. Sorenson et al., 2006). The structure and characteristics of clusters and regional networks are explored in a number of recent studies (e.g. Cantner and Graf, 2006; Cantner et al., 2008a; Graf and Henning, 2008; Ter Wal, 2008). Apparently, when applied to a local or regional network, there is a trade-off as the information is gained at the cost of other valuable aspects of the regional innovation system. In the case of patent data, the observed relations will only include actors that produce patentable innovations, services are clearly underrepresented and important actors for regional development, such as venture capitalists or local government, are not included.

A different approach in terms of data collection is taken in a variety of studies on clusters and industrial districts (e.g. Boschma and Ter Wal, 2007; Giuliani, 2007; Giuliani and Bell, 2005). Here, smaller groups of relevant actors are asked about different types of relations to co-located actors, but because of the high efforts related to this research approach, these studies are – with few exceptions (e.g. Giuliani and Bell, 2008) – static analyses.

Nevertheless, these studies provide us with insights that are to be seen as complementary to existing studies on regional innovation systems and move empirical research in closer harmony with the requirements of an evolutionary economic geography (e.g. Boschma and Martin, 2007; Glückler, 2007).

In his conceptual paper, Glückler (2007) opens a wide field for empirical research by raising important questions about mechanisms of selection, retention, and variation of ties in geographical networks and discussing their consequences. From the number of questions, we attempt to answer those that are related to selection mechanisms at the node and tie level and the consequences for properties of the local network. Our central concern is about how and through which processes the innovation system is affected by the network of interorganizational and interpersonal relationships. But it is not just the relations between existing actors that shape the system, rather, we propose that a great share of the dynamics in such a system stems from the entry and exit of actors and how they relate to each other and integrate themselves into the existing network. What we can observe then is a changing composition of the network that is related to the formation and cutting of ties. Another important aspect of a regional network lies in its capacity to absorb external knowledge (e.g. Bathelt et al., 2004; Graf, 2007). Connecting questions about the importance of internal density and the need for extra-local linkages, we search for the prevalence of both types of relations, how they change, and which actors drive this development.

We attempt to answer these questions by providing case study-based insights into the dynamic, developmental or evolutionary pattern of regional innovation systems. For that, we conduct an analysis of the development of the innovation network of Jena between 1995 and 2001. While at this point we cannot provide a coherent theory of the evolution of networks, we still try to base our analysis on some theoretical building blocks that we consider central for such a theory in the future in section 2. Section 3 introduces the case of Jena and our database, while in section 4 we present our empirical results. The final section 5 puts those results into the perspective of an evolutionary approach towards the development of a regional innovation system.

2. Theoretical building blocks

Innovator networks are the result of interaction between various agents who transfer and exchange knowledge, expertise and ideas among each other in order to create new solutions for technological and economic problems. Hence, those networks are built up by actors collectively engaged in the creation of new ideas they can economize on. There are various aspects characterizing the structure of those networks such as the number and type of actors involved, the number and types of relationships among them, the type of knowledge created, and network measures such as size, density, centralization, or small world properties. An extension of this static view of a network's structure towards a dynamic analysis clearly needs an understanding of actors' decisions to form and cut relations to explain how such systems evolve.

Heterogeneity and the decision to exchange knowledge

Why do actors look for partners to exchange knowledge, expertise and ideas? Why do actors often engage in collective processes of invention and innovation instead of doing it alone? In a rather general way, one line of argument just refers to reducing risk and sharing R&D costs (Baum et al., 2000; Deeds and Hill, 1996). Another line claims that the reason is they are combining complementary assets in order to enhance the propensity of a successful development project (Nooteboom, 1999; Teece, 1986). A third reasoning highlights the internalization of knowledge spillovers (Griliches, 1992). Obviously and rather generally, all three motives overlap and apply simultaneously.

To understand innovator systems, however, the second and the third line of argumentation stand out. Both imply that for generating new ideas different pieces of knowledge have to be combined and those pieces are by no means under the control of a single individual actor. Actors are rather different in their command of technological knowledge and expertise. This just leads us to the basic argument of the resource-based (Barney, 1991; Penrose, 1959) and the dynamic capabilities view (Teece, 1986) of the firm, which we here simply extend beyond the very notion of a firm also to other actors such as research institutions.

The resource-based view of the firm applied to our context suggests an actor should be considered as an ensemble of specific resources that are unique, valuable and difficult to imitate (Barney, 1991). Technological knowledge and technological capabilities are considered to be of that type since they are accumulated over time, show path-dependency and by that are unique to the actor (Teece, 1986). This provides an actor with a certain technology-based competitive advantage that may allow him or her to reap higher benefits or profits. However, they also may constrain an actor in pursuing further progress because for that, additional or different knowledge and competences are required.

The constraining property of knowledge resources in the above sense may force problem solving actors to extend their knowledge base and their repertoire of capabilities. One way to achieve that would be investment in their own learning and R&D. However, as such activities seem to be characterized by cumulativeness and path dependency, they may finally lead to a technological lock-in. In such a situation, an actor's accumulated knowledge and competences may not be suitable anymore to solve current problems.

Hence, it may be in the interest of that actor to gain access to knowledge and competences provided by other actors, may that be firms or public institutions. Besides this, cost considerations may play a role in the sense that it may be less costly to acquire external knowledge, rather than building it up by themselves. For that to work, formalized R&D cooperative arrangements are one way, another way is the informal exchange of knowledge. In both, the exchange of knowledge and competences is ruled either by a formal contract or by an informal agreement governed by trust, fairness, and reciprocity (Schrader, 1991; von Hippel, 1987).

In view of the advantages of working together by exchanging knowledge, expertise and ideas a further question refers to the conditions under which appropriate partners can be found and what kind of 'distances' between heterogeneous actors have to be overcome. In this context, Boschma (2005) distinguishes dimensions of cognitive, social, geographical, institutional, and organizational proximity. The cognitive dimension points to proximity between partners in terms of knowledge and capabilities that is often called technological proximity or technological relatedness. Quite obviously, the relationship here is non-linear in the sense that a too close proximity (not much knowledge can be exchanged) as well as a too large distance (not much of the other's knowledge can be understood) between the partners may not allow technical problems to be solved commonly (Cantner and Graf, 2004, 2006; Cantner and Meder, 2007; Mowery et al., 1998). Here, the concept of absorptive capacities of the recipient of knowledge gains importance (Cohen and Levinthal, 1990) in the sense that those capacities are negatively related to the distance between actors in knowledge space. Furthermore, to find and trust an appropriate cooperation partner, geographical proximity (the other is located in not too far distance) as well as social proximity (the other is known or can easily be approached) are often helpful, especially when we think of informal exchange or the transfer of so-called tacit knowledge (Boschma, 2005; Breschi and Lissoni, 2003; Cantner and Meder, 2008; Singh, 2005). Finally, and referring to factors accompanying and enabling knowledge exchange, institutional proximity refers to common habits, customs, and language, whereas organizational proximity relates to the mode of exchange distinguishing between market transactions, hierarchical designs, and network relations.

Besides these proximity concepts, an efficient and durable exchange of knowledge requires a certain degree of complementarity of partners' knowledge stocks (which to a certain degree is related to technological or cognitive proximity) and based on that reciprocity (not necessarily uno actu) in exchange. Additionally and obviously related to social proximity, for knowledge exchange between actors to be repeated and/or continued, factors such as success, trust, and learning to cooperate become relevant (Gulati, 1999).

Knowledge dynamics and network development

Based on these building blocks it is possible to draw a picture of an evolving network of innovators. The main driving mechanisms are the generation of new knowledge on the one hand and the levelling of knowledge stocks among the actors on the other. The latter is related to the exchange of knowledge, expertise and ideas between partners leading to an increasing similarity of partners' knowledge stocks. The former mechanism has to do with the main objective of cooperating in research, namely the generation of new knowledge. This new knowledge has two effects: first, it advances the partners' knowledge stocks so that they become more similar; second, their knowledge stocks become more dissimilar compared to the knowledge stocks of actors not involved in this partnership. Hence, looking at a group of innovative actors, the exchange of knowledge and expertise leads to more homogeneous knowledge stocks whereas the (local) creation of new knowledge tends to increase the respective heterogeneity among the network actors.

Let us now relate these two mechanisms to the dynamics of an innovator network that for now we assume to be spatially bounded. The expected dynamics depend on the net effect of processes that lead to more heterogeneous individual knowledge stocks and of processes that lead to more homogeneous knowledge stocks. With levelling factors dominating, it will become more difficult for actors to find appropriate partners within the network as the potentials to fruitfully exchange knowledge diminish. For the structural pattern this implies the number of connections among actors will fall and/or actors will exit the network. Actors remaining in the network are presumed to frequently change partners; in this respect, we find that the likelihood of two actors cooperating in period t is negatively affected by cooperating in period $t - 1$ (Cantner and Graf, 2006). Contrariwise, if the dynamics lead to more distinctive individual knowledge stocks, the potentials to exchange knowledge increase. Hence, it becomes more attractive to exchange knowledge and therefore we would expect the number of connections among actors and/or the number of actors to increase as well. As the knowledge stocks of connected actors become more similar, we would expect them to cut existing ties in favour of new, more fruitful relationships.

These bare bones of network dynamics can now easily be extended by a geographical or regional dimension. Assuming that there are clear-cut regional boundaries to an innovator network, one may now ask to what extent and for what reason a network may increase its geographical reach; that is, local actors form extra-local linkages. Two main mechanisms may be at work. The first one has to do with the degree of new knowledge generation within the network. The higher this degree, the more extra-local linkages may be formed as external actors are attracted. As explained above, in such a situation we would also expect an increased number of internal relationships. If, contrariwise, the levelling effect within the network dominates, actors faced by a lower potential for internal partners may look for external relationships, or leave the network and exit. This, then, should go hand in hand with a lower number of internal relations.

The dynamic analysis presented so far can, of course, be extended to other dimensions stated above, complementarity, reciprocity and trust. For example, the levelling effect may also show up as a specialization effect resulting in a lack of complementarity of the knowledge bases. For that, Cantner and Graf (2004) find that the more a regional network is specialized technologically, the more external relationships to exchange knowledge are observed. Last but not least, a lack (abundance) of reciprocity as well as of trust may lead to a lower (higher) density of relations as well as to a reduced (prolonged) duration of the collaboration.

For the present case study, we attempt to identify some of the dynamic patterns introduced. We start with observations on the changing composition of the network and of the structure of relationships within the network of innovators pertaining to Jena. Respective changes will then subsequently be related to the role of:

- potentials available for cooperation that may change over time;
- permanent actors who stay within the network during the whole span of observation;
- actors who enter the network presumably expecting advantages because of cooperation;
- actors who exit the network, which may be because of a low degree of integration and hence poor results;

- the geographical dimension, in the sense that to solve technological problems an actor may not find an appropriate cooperation partner nearby (regionally internal partners) but only further away, outside the local system (regionally external partners).

Obviously, to find some explanation for the changing structure of innovation networks, we have to take an actor's perspective and use information on the technological relationships among them. What we have to ignore or can only indirectly infer are the motivations and further determinants that lead an actor to engage in network cooperation.

3. The case and the data

The above five dimensions of a developing innovation system are analysed for the case of the local innovation network of Jena between 1995 and 2001. Hence, we do not look at the local innovation system of Jena but at a subset of this system, the network of innovators. Thus, we leave out political actors, norms and institutions that are considered as additional core actors and elements of a local innovation system.[2] In order to track the development of the Jena innovators network, we distinguish two sub-periods, 1995–97 labelled P1 and 1999–2001 labelled P2. We leave out 1998 in order to have the two sub periods of equal length of three years and to have a clear separation between them.

We construct the network of innovators by using information from applications at the German Patent Office, and select those with at least one inventor residing in Jena. For the period under consideration, we take into account 1114 patent applications, with 334 applicants and 1827 inventors, of which 977 are located outside Jena. We distinguish between the technological potential to cooperate, the kind of relationships of the cooperating actors, and the geographical dimension of the relationships. From the patent information we have at hand, these relationships or dimensions can be constructed in the following way. First, there is potential to cooperate between actors if they apply within the same technology class. Second, if for a specific patent we find more than one applicant (co-application), then we consider this as cooperation (*co*) between those applicants. Third, if we find patents by distinct applicants on which the same inventor is stated, we consider the applicants of these patents to be connected via scientist mobility (*sm*) – seemingly, this inventor worked for all the applicants.[3] Fourth, the regional dimension is taken into account by looking at the location of the applicants. In the case of a relationship between applicants located in Jena, the respective knowledge flows are internally oriented. An external orientation of a relationship shows up when a local actor is linked to an actor located outside Jena.

4. Results

Actor development

Table 17.1 gives a first view on the development of the network in terms of the actors involved. The network size increased as the total number of innovators jumped from 139 in the first period to 189 in the second period; 32 innovators were members of the network in both periods, 107 actors exited between P1 and P2, and 157 entered. Within

Table 17.1 Types of actor in the innovation network

		Internal	External	Total	Share of external (%)
P1	Exiting	49	58	107	54.2
	Permanent	20	12	32	37.5
	Total	**69**	**70**	**139**	
P2	Entering	66	91	157	58.0
	Permanent	20	12	32	37.5
	Total	**86**	**103**	**189**	

the group of permanent innovators, 37.5 per cent were not located in Jena but were related to actors in Jena in both periods. The share of external partners that exited or entered the network was greater than 50 per cent in both cases, with a slightly higher percentage in the case of entering actors. This could be interpreted as a slight tendency of the Jena innovator network to become more externally oriented.

The network of innovators and its change over time

In the next step, we look at relationships between the innovators via common inventors; that is, the network of personal relationships (*pr*). Herein, we combine the two types, cooperation *co* and scientist mobility *sm* – as defined above. Figure 17.1 shows the *pr*-network, where we distinguish the *co*-network with black edges and the *sm*-network with grey edges.

In Table 17.2, we report the statistics for the networks of cooperation (co_t), scientist mobility (sm_t), and the (aggregate) network of personal relationships (pr_t). From the various indicators and their change we can infer how the connectedness of the three networks develops.[4]

We start with the component structure. A component is a sub-structure of the network, where all members of the component can reach each other, while there is no connection to members of other components. Since we analyse relations that measure the flow of information and knowledge, a high share of actors in the largest component (the one presented in Figure 17.1) means that many actors within the Jena region can exchange knowledge and by that build up something like a common knowledge base. The analysis of the component structure shows a trend towards decreasing fragmentation. In addition to the fact that the network is growing, the share of innovators in the main component increases from 42.4 per cent to 49.7 per cent in the *pr*-network, from 8.6 per cent to 31.2 per cent in the *co*-network, and from 25.2 per cent to 32.3 per cent in the *sm*-network. The share of isolates in the *pr*-network stays roughly constant with a decreasing tendency in the *sm*-network and an increasing one in the *co*-network. Hence, this increasing connectedness of the actors is a first indication of intensified knowledge flows within the Jena network.

Density is another measure that provides some information about the intensity of knowledge flows. If g is the size of the network as measured by the number of actors and d_i is the degree, that is, the number of connections, of actor i, $i = 1, \cdots, g$, then the density D of the network is defined as the number of all linkages divided by the number of possible linkages within the network $D = \Sigma_{i=1}^{g} d_i/(g^2 - g)$. Looking at that indicator

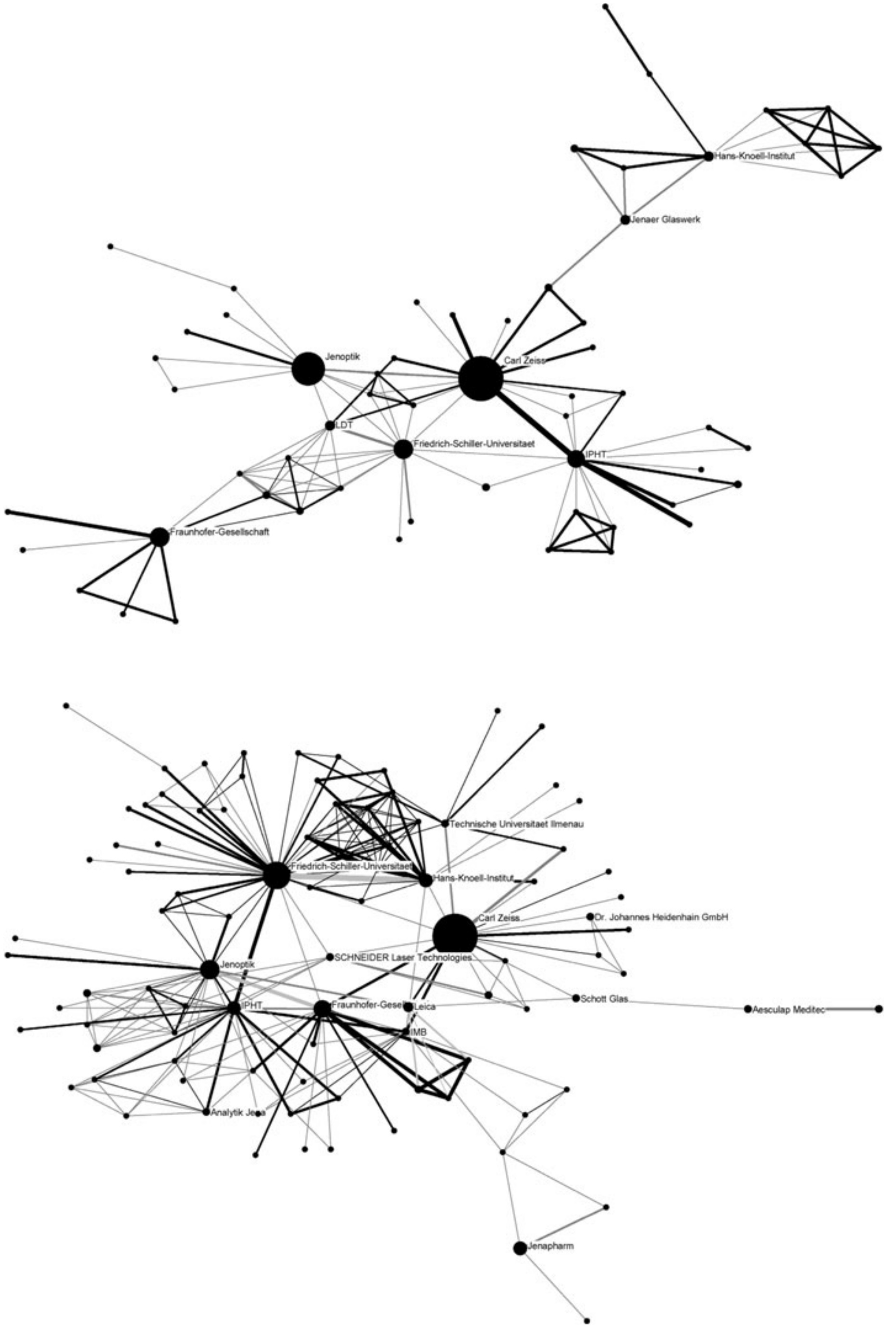


Figure 17.1 *Innovator network in P1 (top) and P2 (bottom) – cooperations (black) and scientist mobility (grey)*

Table 17.2 Network statistics related to the pr-, co- and sm-networks

	pr_{P1}	pr_{P2}	co_{P1}	co_{P2}	sm_{P1}	sm_{P2}
No. of actors	139	189	139	189	139	189
Share of largest component	42.4%	49.7%	8.6%	31.2%	25.2%	32.3%
Share of isolates	30.2%	30.7%	51.1%	57.7%	54.7%	55.0%
Density	0.040	0.037	0.029	0.027	0.010	0.010
Mean degree[a]	2.201	2.815	1.065	1.543	1.137	1.344
Network centralization[a]	0.109	0.184	0.051	0.137	0.102	0.057
Clustering coefficient	2.452	2.191	3.634	2.833	0.856	0.648
Average distance[b]	3.581	2.799	1.699	2.974	2.634	3.325

Note: [a] Networks have been dichotomized; [b] among reachable pairs.

we find a decrease in the *co*-network (0.029 to 0.027) and a rather constant value for the *sm*-network (0.010). The overall effect is dominated by the effects of cooperation, which leads to a *pr*-network that is less dense in the second period (0.040 to 0.037). This result suggests that the relative intensity of knowledge flows decreased.

However, the density measure is not invariant to the size of the network. We therefore look at the mean degree as another measure of connectedness, which is defined as $md = \Sigma_{i=1}^{g} d_i / g$. It states how many connections each actor on the average has. Here we find an increasing connectedness in all three networks and thus an indication for an increased flow of knowledge and information among the actors.

Asking how this increased connectedness is distributed over the network, we can look at the centralization index. The degree centrality of actor *i* is the number of its ties divided by the number of possible ties $C_i = d_i / (g - 1)$. The network centralization is then given by $C = \Sigma_{i=1}^{g} (\max(C_i) - C_i) / (g - 2)$. For the *pr*-network, centralization increases from 0.109 to 0.184, which implies an increased importance of central actors for the flow of knowledge in the whole network. Looking at the other two networks, we observe a sharp increase in the *co*-network (from 0.051 to 0.137) and a decrease in the *sm*-network (from 0.102 to 0.057). Connections through cooperation are therefore increasingly established among or with the most central actors, while knowledge flows through mobility are more often found between more peripheral actors.

Another structural measure for a network is the overall clustering coefficient, where an increase hints towards local coherence, that is, intensified grouping of actors in densely connected clusters. This measure is calculated by averaging the clustering coefficients of all actors within the network. These node-level clustering coefficients are calculated as the density of the neighbourhood, that is, the network of actors directly linked to the respective actor. For all three networks we find less clustering in the second compared to the first period and thus less intense grouping of the actors. The average distance between actors measures the ease of knowledge flows and is often related to the rate of knowledge diffusion (Cowan and Jonard, 2004). For the *co*- and the *sm*-network, we find an increase whereas for the total *pr*-network this measure decreases. This seems to indicate that there is a complementary relation between knowledge flows via cooperation and via scientist mobility.

Table 17.3 Network statistics for the networks of technological relatedness

	$tech_{P1}$	$tech_{P2}$
No. of actors	139	189
Share of largest component	97.1%	94.7%
Share of isolates	1.4%	0.5%
Density	0.149	0.165
Mean degree[a]	19.165	28.730
Network centralization	0.602	0.717
Overall graph clustering coefficient	1.228	1.200
Average distance (among reachable pairs)	2.003	1.879

Note: [a] Networks have been dichotomized.

Summarizing these results for the Jena region, we find decreasing fragmentation combined with a higher degree of connectedness where the large, core actors within the network increasingly focus on formal cooperation, while the smaller, surrounding or peripheral actors rather intensify their contacts through informal, personal relations. In the following, we want to analyse more closely the determinants of this increasing connectedness of the Jena network of innovators that took place despite an increase in the total number of actors. We start with the analysis of the changing potentials to cooperate.

Potential to cooperate

As noted above, technological relatedness is one of the preconditions for actors to cooperate and to exchange knowledge. Any increase in this relatedness may be a cause for personal relations to become more frequent and for the whole system to become more interconnected. In this sense, technological relatedness or proximity can be interpreted as the potential to cooperate.

Innovators are assumed to be close in technology space if they have patented in the same technology classes. For the technological classes, we use a concordance between the IPC code of the patent classification and 30 technologies that has been elaborated jointly by the Fraunhofer-Institut für Systemtechnik und Innovationsforschung (FhG-ISI), the Observatoire de Sciences et des Techniques (OST), and the Science and Technology Research Policy Unit of the University of Sussex (SPRU). The more classes two innovators have in common, the higher is their degree of technological overlap and thus their technological relatedness. These technological relations are used to build up a network of technological overlap that we interpret as the potential to cooperate.

Table 17.3 shows the statistics with respect to this network. Obviously the cohesion of the network increases – according to the measures density and mean degree – with the number of actors from P1 to P2. Thus the potential to cooperate increased over time.

Since we consider a very broad concept of technological overlap with only 30 technological classes, in both periods, the networks turns out to be highly connected. For each we observe roughly 95 per cent of the actors being part of the largest component. The potential network has become increasingly connected as evidenced by the measures of density (0.149 to 0.165) and also by an increase of the mean degree from 19.165 to

28.730. We also find an increase in the centralization of the network from 0.602 to 0.717, which means that peripheral actors in the network are more strongly connected to actors in the centre and/or less connected to other peripheral actors. The overall clustering coefficient slightly decreases from 1.228 in the first period to 1.200 in the second period. At the same time, the average distance between actors decreases from 2.003 to 1.879.

Summarizing, we find increasing cohesion, which is interpreted as a stronger focus on core competencies or technologies, where the fields of activity of the central actors become increasingly important for the whole network. With respect to the increased connectedness of the *pr*-network we presume that the increasing potential to cooperate goes hand in hand with this change over time. Before we go on to analyse this relation in more detail, let us dwell a little bit more on the development of technological overlap and consider actors entering and exiting the network.

As we recall from Table 17.1, the number of actors changes between P1 and P2, with only 32 actors being part of the network in both periods. We thus ask whether the observed increasing technological cohesion can be related to the following three groups of actors – entering, exiting, and permanent innovators. Calculating mean degrees for these different groups, we distinguish between relations within the group and between the groups. The resulting values and tests for the significance of differences between the mean degrees are given in Table 17.4. Since we cannot assume a normal behaviour of the degrees of the network members, we perform the nonparametric Mann-Whitney or Wilcoxon rank sum test to examine our above reasoning. The tests are performed by column, that is, we compare exiting innovators of the first period with the entering innovators of the second period and so on.

Regarding the different roles of exiting innovators and entrants, we observe a stronger

Table 17.4 Technological overlap: mean degree (within and between blocks)

	Within	Between	Between	Within
P1	Exit	Exit and permanent	Permanent and exit	Permanent
N	107	107	32	32
Mean degree	8.729	6.738	22.531	15.125
Standard deviation	(5.984)	(4.187)	(22.361)	(14.914)
P2	Entry	Entry and permanent	Permanent and entry	Permanent
N	157	157	32	32
Mean degree	18.191	7.592	37.250	19.563
Standard deviation	(16.196)	(4.825)	(38.756)	(17.629)
Significance of difference between mean degrees				
	Mann-Whitney		Wilcoxon rank sum test	
W	5854	7906.5	101.5	130
p-value	0.000	0.208	0.002	0.011

Note: One-sided tests are performed with H_0 as no difference between samples and H_1 in the direction of the observed differences.

connectivity within the entering group itself (in the first column 18.191 compared to 8.729 for the exiting group), but also with respect to the linkages with the permanent group (in the second column 7.592 compared to 6.738 for exit). While the former observation, that entrants are a more homogeneous group, is significant, the latter is not at an acceptable level of significance (*p*-value of 0.208). The values in the second and third column are based on the same number of connections but in calculating the mean they are divided by a different number of focal actors. Accordingly, we notice more technological connections between the permanent innovators and the entering innovators (37.250) as compared to the exiting group (22.531). The permanent innovators themselves increase their technological overlap significantly between the two periods from 15.125 to 19.563 (fourth column).

With respect to our previous result of an increasing cohesion of the network of technological overlap, the findings here indicate that both the permanent innovators as well as the entrants (compared to the exiting innovators) tend to increasingly concentrate on the technological core competencies of the network as a whole. And with respect to the increasing connectedness of the *pr*-network, we presume that the permanent actors' degree of cooperation should have increased between P1 and P2, and the degree of cooperation with entrants should have been higher than with exiters.

Entry and exit to the network

In order to develop further on this presumption, we compute mean degrees for several groups of actors in both periods with respect to personal relationships. We again distinguish between relationships based on cooperations (*co*) and on scientist mobility (*sm*). The respective results are stated in Tables 17.5 and 17.6.

Table 17.5 Cooperation: mean degree (within and between blocks)

	Within	Between	Between	Within
P1	Exit	Exit and permanent	Permanent and exit	Permanent
N	107	107	32	32
Mean degree	3.084	0.710	2.375	2.563
Standard deviation	(5.207)	(1.873)	(5.375)	(5.346)
P2	Entry	Entry and permanent	Permanent and entry	Permanent
N	157	157	32	32
Mean degree	2.242	1.516	7.438	3.938
Standard deviation	(4.424)	(2.623)	(17.005)	(6.710)
Significance of difference between mean degrees				
	Mann-Whitney		Wilcoxon rank sum test	
W	9191.5	7111	34.5	49
p-value	0.066	0.003	0.136	0.100

Note: One-sided tests are performed with H_0 as no difference between samples and H_1 in the direction of the observed differences.

Table 17.6 Scientist mobility: mean degree (within and between blocks)

	Within	Between	Between	Within
P1	Exit	Exit and permanent	Permanent and exit	Permanent
N	107	107	32	32
Mean degree	0.561	0.514	1.719	0.938
Standard deviation	(1.361)	(0.883)	(2.976)	(1.900)
P2	Entry	Entry and permanent	Permanent and entry	Permanent
N	157	157	32	32
Mean degree	0.497	0.637	3.125	2.500
Standard deviation	(0.965)	(1.415)	(4.030)	(2.700)
Significance of difference between mean degrees				
	Mann-Whitney		Wilcoxon rank sum test	
W	8166.5	8613	66.5	20
p-value	0.695	0.668	0.008	0.000

Note: One-sided tests are performed with H_0 as no difference between samples and H_1 in the direction of the observed differences.

In the last column of both tables the permanent innovators themselves (fourth column in Tables 17.5 and 17.6) show a significant increase in linkages through cooperation (from 2.563 to 3.938) and especially in scientist mobility (from 0.938 to 2.500). Thus, parallel to the increased technological overlap in this group, the connectedness among the incumbents becomes more intense over time.

The first column in both tables regards the comparison of the connectedness between exiting and entering innovators within their respective group. The exiting innovators show significantly more linkages through cooperation than the entering group (3.084 versus 2.242) and slightly, but not significantly more linkages through scientist mobility (0.561 versus 0.497). The second and third columns report the mean degree between permanent and exiting innovators in P1 and between permanent and entering innovators in P2. The two columns differ with respect to the group of actors for which this index is computed, for the exiting or entering innovators in column 2 and for the permanents in column 3. For the *co*-network in column 2, we find that entering innovators cooperate significantly more with the permanent ones than the exiting innovators did (1.516 versus 0.710). With respect to the *sm*-network, entering innovators have also more linkages through scientist mobility (0.637 versus 0.514) even though not to a significant degree.

Summarizing these results, innovative entrants in Jena seem to be better integrated into the network of personal relations than actors who, for whatever reasons, stopped innovating. This finding is consistent with results of Powell et al. (1999) that the network position has an important influence on firm performance. The observations regarding the connectedness of entering and exiting innovators within their group are not contradictive to our argument. Actors that enter such a network are certainly more aware of incumbents than of actors who enter during the same period. Information about other potential partners for knowledge exchange needs time to develop.

Regional dimension

Another question, regarding increasing connectedness of the Jena innovator network, is related to the issue of whether this is a result of an increase in internal relationships or whether more external partners are responsible for this development. In the former case, this could be on indication of the danger of getting into a technological lock-in situation (Grabher, 1993). In the latter case, the external pipelines may either be interpreted as protection against such an unfortunate situation (see Bathelt et al., 2004) or as an indicator of the attractiveness towards external partners. From Table 17.1 we take that the share of external actors increased slightly from about 54 per cent in P1 to 58 per cent in P2, suggesting that knowledge exchange with external partners tends to intensify over time.

In Table 17.7 we present a more fine-grained analysis of these external relations. We distinguish relations between only internal actors (*internal*) on the one hand and between an internal and an external partner (*external*) on the other. The development of these two relations from P1 to P2 as well as the respective development of the ratio between the two is stated for both types of relation (*co* and *sm*) and for different types of actor (private vs. public, persistent vs. non-persistent). For the *co*-network we find that both *external* and *internal* increase in absolute numbers over time. However, the increase in *internal* is comparatively larger, leading to a decrease in the ratio *external/internal*. Thus, the development of the *co*-network clearly shows an increasing inward orientation in relative terms (from 1.682 to 1.563), although in absolute terms the number of external actors in the network has been increasing. The same holds true for the *sm*-network (from 2.000 to 1.605). We conclude from this pronounced inward orientation combined with an increased number of external actors that the knowledge portfolio and cooperation potential of Jena seems to be quite attractive.

In view of these results, in a further step we distinguish between private and public actors to see if certain types of actor are responsible for the observed development. The results in Table 17.7 indicate that while in both groups we observe an increasing inward orientation (the share *external/internal* always decreases), this tendency seems much more pronounced for the public actors compared to private actors – compare the decline for public actors in the *co*-network (*sm*-network) from 1.857 to 1.524 (from 1.769 to 1.273) for the public actors to the decline from 1.649 to 1.593 (from 2.091 to 1.741) for the private actors.

Alternatively, if we distinguish between permanent innovators and the ones that exit or enter, we observe that for these groups, the development goes in different directions. First, we find that the entrants are on average more outward oriented than the actors that exited the network; hence in Table 17.7 the non-permanent actors are responsible for an increasing outward orientation regardless of the type of relation (from 1.929 to 2.140 in the *co*-network and from 1.958 to 2.367 in the *sm*-network). To explore the manifold reasons for this behaviour surely lies beyond the scope of our exercise and would require in-depth interviews with the responsible actors. All we can offer at this point is an informed guess and plausible interpretation. In the tradition of industry in the GDR (the former East Germany), economic and technological activities in Jena were organized in

Table 17.7 Internal and external relations in the network

	co		*sm*	
	P1	P2	P1	P2
All				
External	74	150	92	122
Internal	44	96	46	76
External/internal	1.682	1.563	2.000	1.605
Private				
External	61	86	69	94
Internal	37	54	33	54
External/internal	1.649	1.593	2.091	1.741
Public				
External	13	64	23	28
Internal	7	42	13	22
External/internal	1.857	1.524	1.769	1.273
Non-permanent				
External	54	107	47	71
Internal	28	50	24	30
External/internal	1.929	2.140	1.958	2.367
Permanent				
External	20	43	45	51
Internal	16	46	22	46
External/internal	1.250	0.935	2.045	1.109

the large Carl-Zeiss combine. During the transformation process after German unification, this combine was split and many workers were laid off at the beginning of the 1990s. This generation of workers was socialized in the Carl-Zeiss combine and newly founded firms recruited workers who had close contacts with each other because of their common history and past collaboration. This common heritage is of decreasing importance for firms that were founded later on and for which local labour was not as abundantly available as before. With this in mind, an increasing outward orientation of the entering actors has to be seen as a natural development of a system in transformation.

Second, the permanent innovators increasingly focus on internal relations; Table 17.7 shows that this is especially true for linkages through mobility, where internal relations more than double from P1 to P2. One of the reasons for this development might be initiatives to foster academic start-ups and a culture of entrepreneurship in general during that period. Scientists and engineers who have been working for large employers in the region and founded new firms or work in these new establishments could account for that increased mobility. In addition, for incumbents the new establishments provide increasing possibilities for knowledge exchange not only as the system grows, but also as the technological orientation of these new actors is more in line with the competencies of incumbents.

Summarizing these results we find that even though the share of external actors in the Jena network of innovators increased slightly over time, linkages within the region enlarged their share. This tendency is especially true for public actors, but applies to private ones as well. Distinguishing actors according to their presence in the network, we find a development in opposing directions. Permanent actors are increasingly inward oriented, while for non-persistent actors we find that entrants are more outward oriented than the exiting innovators. The antenna function that permanent and central actors – especially universities – are assumed to fulfil, seemingly shows diminishing importance, at least in relative terms.

5. Conclusion

In this chapter we discuss some issues pertaining to the development or evolution of innovation networks. Actors, heterogeneous in terms of technological knowledge and competences, who seek to cooperate do so to gain access to other actors' knowledge and exchange it for their own knowledge. Connecting such cooperative actors within one region, one ends up with a system or network of innovating firms. Consisting of bilateral knowledge relationships, the network of innovators may change its size and its structure over time, new relationships come into existence, existing relationships may be cancelled, new actors join the system, and other actors leave it. Based on those structure-generating factors one can presume that the system as a whole in some periods tends to become more coherent and the actors to be more connected, whereas in other periods the tendency to disconnect and to weaken the system may prevail.

Our study on the Jena network of innovators represents just the case where the technological potentials for cooperation increased over time accompanied by a growth in the number of actors involved and a higher connectedness among them. These tendencies are nicely related to the specific contributions of permanent, entering, and exiting actors, as well as of actors internal to the region of Jena and those who are external. Exiters tend to be less integrated into the innovators' network than entrants. Although the share of external actors increased over time the share of external linkages diminishes over time. Thus we observe an increasing orientation of the network's innovation activities with internal partners. The fact that permanent actors strongly intensify their internal relations seems to be the main driver of this development. This pattern can be interpreted as documenting the attractiveness of the Jena knowledge pool, for internal as well as external actors.

What we have presented here is the study of a single innovator network and its change over time. Such a network has to be viewed as a constituent part of an innovation system. One of the drivers of innovations, institutions, cannot be observed directly within this methodology, as we can only observe the relations but not the rules that govern them. Evolutionary economic geography is surely more than what we present here. It is not about single places but it is about the patterns of economic activity in space. We tried to account for this aspect by not only focusing on the changing relations within the regional network, but also by analysing the relations to actors located in other regions. A theory of a networked economy has to encompass these different types of relations, as becomes apparent in the notions of local buzz and global pipelines (Bathelt et al., 2004). As such, an economy can be viewed as a small world network not only in terms of relations (Cowan and Jonard, 2004), but also in terms of geography. A densely connected network as in our case of Jena is linked to global knowledge channels via their external relations. It is of critical importance for any regional development that these linkages go in both directions (otherwise it would be a brain drain rather than a brain gain) and that they help to improve the regional knowledge base. For a thorough judgement of our observations, these qualitative aspects of internal and external relations are crucial.

Notes

1. For an overview see Edquist (2004).
2. A detailed picture of the Jena innovation system is given in Graf (2006) and the role of intermediation and political actors within the system is analysed in Cantner et al. (2008b) and Cantner et al. (2009).
3. For an overview of methods for creating networks from patent information see Breschi and Lissoni (2004). For a more detailed description of our specific application see Cantner and Graf (2006).
4. For further details on the definition and calculations of the various network measures, please refer to Wassermann and Faust (1994).

References

Allen, R.C. (1983), 'Collective invention', *Journal of Economic Behavior and Organization*, **4**, 1–24.
Barabasi, A., H. Jeong, Z. Neda, E. Ravasz, A. Schubert and T. Vicsek (2002), 'Evolution of the social network of scientific collaborations', *Physica*, **A 311**, 590–614.
Barney, J. (1991), 'Firm resources and sustained competitive advantage', *Journal of Management*, **17**, 99–120.
Bathelt, H., A. Malmberg and P. Maskell (2004), 'Clusters and knowledge: local buzz, global pipelines and the process of knowledge creation', *Progress in Human Geography*, **28**, 31–56.
Baum, J., T. Calabrese and B. Silverman (2000), 'Don't go it alone: alliance network composition and start-ups' performance in Canadian biotechnology', *Strategic Management Journal*, **21**, 267–94.
Boschma, R. (2005), 'Proximity and innovation: a critical assessment', *Regional Studies*, **39**, 61–74.
Boschma, R. and R. Martin (2007), 'Editorial: Constructing an evolutionary economic geography', *Journal of Economic Geography*, **7**, 537–48.
Boschma, R. and A.L.J. Ter Wal (2007), 'Knowledge networks and innovative performance in an industrial district: the case of a footwear district in the south of Italy', *Industry and Innovation*, **14**, 177–99.
Braczyk, H., P. Cooke and M. Heidenreich (1998), *Regional Innovation Systems: The Role of Governances in a Globalized World*, London: UCL Press.
Breschi, S. and F. Lissoni (2003), 'Mobility of skilled workers and co-invention networks: an anatomy of localized knowledge flows', *Journal of Economic Geography*, **9** (4), 439–468.
Breschi, S. and F. Lissoni (2004), 'Knowledge networks from patent data: methodological issues and research targets', CESPRI Working Paper, No. 150.
Cantner, U. and Graf, H. (2004), 'Cooperation and specialization in German technology regions', *Journal of Evolutionary Economics*, **14**, 543–62.
Cantner, U. and H. Graf (2006), 'The network of innovators in Jena: an application of social network analysis', *Research Policy*, **35**, 463–80.
Cantner, U. and A. Meder (2007), 'Technological proximity and the choice of cooperation partner', *Journal of Economic Interaction and Coordination*, **2**, 45–65.
Cantner, U. and A. Meder (2008), 'Innovators and the diversity of innovation systems', *Applied Economics Quarterly Supplement*, **54** (59), 9–26.
Cantner U., E. Conti and A. Meder (2009), 'Networks and innovation: the role of social assets in explaining firms' innovative capacity – evidence from the Jena region', Jena Economic Research Papers, 040–2009 (forthcoming in *European Planning Studies*).
Cantner, U., A. Meder and A.L.J. Ter Wal (2008), 'Innovator networks and regional knowledge base', Jena Economic Research Papers, 2008–042.
Cantner, U., A. Meder and T. Wolf (2008), 'Intermediation, reciprocity and compatibility in regional innovation systems – an interregional comparison', Jena Economic Research Papers, 2008–081.
Cohen, W.M. and D.A. Levinthal (1990), 'Absorptive capacity: a new perspective on learning and innovation', *Administrative Science Quarterly*, **35**, 128–52.
Cooke, P. and K. Morgan (1994), 'The regional innovation system in Baden-Wurttemberg', *International Journal of Technology Management*, **9**, 394–429.
Cowan, R. and N. Jonard (2004), 'Network structure and the diffusion of knowledge', *Journal of Economic Dynamics and Control*, **28**, 1557–75.
Deeds, D. and C. Hill (1996), 'Strategic alliances and the rate of new product development: an empirical study of entrepreneurial biotechnology firms', *Journal of Business Venturing*, **11**, 41–55.
Edquist, C. (2004), 'Systems of innovation – perspectives and challenges', in J. Fagerberg, D.C. Mowery and R.R. Nelson (eds), *Oxford Handbook of Innovation*, Oxford: Oxford University Press.
Fleming, L., C. King and A. Juda (2006), 'Small worlds and regional innovation', SSRN eLibrary, 892871.

Giuliani, E. (2007), 'The selective nature of knowledge networks in clusters: evidence from the wine industry', *Journal of Economic Geography*, **7**, 139–68.
Giuliani, E. and M. Bell (2005), 'The micro-determinants of meso-level learning and innovation: evidence from a Chilean wine cluster', *Research Policy*, **34**, 47–68.
Giuliani, E. and M. Bell (2008), 'Industrial clusters and the evolution of their knowledge networks: back again to Chile', SPRU Working Paper, 171.
Glückler, J. (2007), 'Economic geography and the evolution of networks', *Journal of Economic Geography*, **7**, 619–34.
Grabher, G. (1993), 'The weakness of strong ties: the "lock-in" of regional development in the Ruhr area', in G. Grabher (ed.), *The Embedded Firm: On the Socioeconomics of Industrial Networks*, London: Routledge, pp. 255–77.
Graf, H. (2006), *Networks in the Innovation Process: Local and Regional Interactions*, Cheltenham, UK and Northampton, MA, USA: Edward Elgar.
Graf, H. (2010), 'Gatekeepers in regional networks of innovators', *Cambridge Journal of Economics*, doi:10.1093/cje/beq001.
Graf, H. and T. Henning (2008), 'Public research in regional networks of innovators: a comparative study of four East-German regions', *Regional Studies*, **43** (10), 1349–68.
Griliches, Z. (1992), 'The search for R&D spillovers', *Scandinavian Journal of Economics*, **94**, 29–47.
Gulati, R. (1999), 'Network location and learning: the influence of network resources and firm capabilities on alliance formation', *Strategic Management Journal*, **20**, 397–420.
Jaffe, A.B., M. Trajtenberg and R. Henderson (1993), 'Geographic localization of knowledge spillovers as evidenced by patent citations', *Quarterly Journal of Economics*, **108**, 577–98.
Keeble, D., C. Lawson, B. Moore and F. Wilkinson (1999), 'Collective learning processes, networking and "institutional thickness" in the Cambridge region', *Regional Studies*, **33**, 319–32.
Kline, S. and N. Rosenberg (1986), 'An overview of innovation', in R. Landau and N. Rosenberg (eds), *The Positive Sum Strategy*, Washington, DC: National Academics Press, pp. 275–305.
Lundvall, B.A. (1992), *National Systems of Innovation: Towards a Theory of Innovation and Interactive Learning*, London: Pinter Publishers.
Moody, J. (2004), 'The structure of a social science collaboration network', *American Sociological Review*, **69**, 213–38.
Mowery, D.C., J.E. Oxley and B.S. Silverman (1998), 'Technology overlap and interfirm cooperation: implications for the resource-based view of the firm', *Research Policy*, **27**, 507–23.
Nelson, R.R. (ed.) (1993), *National Innovation Systems: A Comparative Analysis*, Oxford: Oxford University Press.
Nooteboom, B. (1999), 'Innovation, learning and industrial organisation', *Cambridge Journal of Economics*, **23**, 127–50.
Penrose, E. (1959), *The Theory of the Growth of the Firm*, London: Wiley & Sons.
Powell, W.W., K.W. Koput, L. Smith-Doerr and J. Owen-Smith (1999), 'Network position and firm performance: organizational returns to collaboration in the biotechnology industry', in S.B. Andrews and D. Knoke (eds), *Research in the Sociology of Organizations: Networks In and Around Organizations*, Westport, JAI Press, pp. 129–59.
Saxenian, A. (1994), *Regional Advantage*, Cambridge, MA: Harvard University Press.
Schrader, S. (1991), 'Informal technology transfer between firms: cooperation through information trading', *Research Policy*, **20**, 153–70.
Singh, J. (2005), 'Collaborative networks as determinants of knowledge diffusion patterns', *Management Science*, **51**, 756–70.
Sorenson, O., J.W. Rivkin and L. Fleming (2006), 'Complexity, networks and knowledge flow', *Research Policy*, **35**, 994–1017.
Teece, D.J. (1986), 'Profiting from technological innovation: implications for integration, collaboration, licensing and public policy', *Research Policy*, **15**, 285–305.
Ter Wal, A.L.J. (2008), 'Cluster emergence and network evolution: a longitudinal analysis of the inventor network in Sophia-Antipolis', Papers in Evolutionary Economic Geography, 08.10, Utrecht, University of Utrecht.
von Hippel, E. (1987), 'Cooperation between rivals: informal know-how trading', *Research Policy*, **16**, 291–302.
Wassermann, S. and K. Faust (1994), *Social Network Analysis: Methods and Applications*, Cambridge: Cambridge University Press.

PART 4

INSTITUTIONS, CO-EVOLUTION AND ECONOMIC GEOGRAPHY

18 An evolutionary approach to localized learning and spatial clustering

Anders Malmberg and Peter Maskell

1. Introduction

Economic geographers approach the spatial aspects of economic development from different angles. They pursue macro analyses of the economic growth of spatial entities such as cities, regions or nations. But they also conduct micro- or meso-level analyses of the development of firms or systems of related firms, such as industries, networks or clusters. In either case, the main concern is often the relation between economic specialization on the one hand, and economic performance on the other.

Thus, macro-level economic geography departs from questions like: Why do some regions (cities, nations) prosper while others don't? Why is there regional specialization – between as well as within countries – and how are such patterns reproduced? Natural resources and demography have an impact, alongside institutional settings, knowledge structures and the general macroeconomic environment. The most influential models dealing with uneven regional development emphasize processes of cumulative causation and the fact that a region that, for one reason or the other, has taken the lead will gradually tend to strengthen its position, partly at the expense of surrounding, less developed regions (Hirschman, 1958; Krugman, 1991; Myrdal, 1957; Ullman, 1958).

Micro- and meso-level analyses, on the other hand, focus on the location of the firm and ask questions like: How is the performance and competitiveness of a firm affected by the conditions that prevail in its immediate environment? Why do similar and related firms tend to agglomerate in certain places? A key factor here is the impact of spatial proximity between firms on various interaction processes (rivalry, collaboration, imitation) that allegedly lead to new knowledge and competitiveness. Supply and skills of local labor, appropriateness of the physical infrastructure and institutional framework are other important attributes of attractive business locations.

Even though space is the constituent aspect of economic geography, the core questions of the discipline are also intimately related to time. Different regions have various capabilities at their disposal – resources, institutions, infrastructures, skills, norms and values – and these make up the basis for their development prospects. The capabilities of a region often date back in history as combinations of chance events, natural endowments, and the effects of previous human activity. Some specific feature in the local environment may determine a firm's choice of location, but the presence of the firm will also affect this environment and influence the cost and quality of the inputs available.

In this sense a geographical location – be it a city, a region or a nation – can be thought of as having a memory that directs the path of subsequent development. Over time, particular choices, themselves framed by past decisions, open up new strands of economic development, but preclude others (Arrow, 1962). The downside tends to be expressed in terms of lock-in to a fated path where development is constrained within a gradually more narrow range of possibilities that ultimately leads to stagnation or decline when confronted with radical technological shifts or wider market changes (Arthur, 1989; Liebowitz and Margolis, 1995).

The rapidly growing literature within evolutionary economic geography (see Boschma and Frenken, 2006 or Martin and Sunley, 2006 for overviews) has added to our understanding of regional economic development, but contributions in this emerging tradition rarely specify the underlying micro-level processes that include individuals or firms with incentives, which, when acted upon, could account for the development under scrutiny (Agassi, 1960; Coleman, 1986, 1990; Frenken and Boschma, 2007).

The objective of this chapter is to demonstrate how processes of knowledge evolution and their institutional underpinnings make up the core of evolutionary economic geography. More specifically the chapter has three aims. The first is to show how micro-level action provides insights needed when investigating evolutionary processes of knowledge creation in a spatial setting. The second aim is to analyze how macro-level institutional dynamics form development paths at the aggregate level of cities, regions or nations. The third aim is to use such micro- and macro-level evolutionary insights to develop an explanatory scheme by which we can review and reinterpret the coming into existence and further evolution of spatial clusters of similar and related firms and industries (Richardson, 1972).

The chapter is structured in five main sections. Next, in section 2, we take a closer look at the drivers, mechanisms and barriers to knowledge creation and acquisition at the micro level. We discuss how cognitive frameworks and spatial settings give rise to mechanisms that guide individual action in general and knowledge creation in particular into specific paths. In the subsequent section we engage with the emergence and possible transformation of territorially specific nuances of general market institutions and other higher-order concepts of localized capabilities and their role for knowledge creation. In section 4, we apply the approach developed in earlier sections on the specific case of clusters of similar and related economic activity. In the final section the argument is summarized and some general conclusions are drawn.

2. Micro-level foundations for knowledge creation: routines within and among firms

The literature on the evolution of knowledge has grown immensely in recent decades.[1] It is now well established how learning from experience, by trial and error or by repetition (Arrow, 1962; Scribner, 1986), gives rise to incremental improvements that can accumulate and gradually result in new and better ways of doing things (Boldrin and Scheinkman, 1988; Thorndike and Rock, 1934). Over time, some novel ways are rejected when the results are confronted with reality, while others function comparatively well and become embedded in the specific routines of individuals or firms (Levitt and March, 1988). Building routines means encoding incentives and constraints of the particular setting (Coriat and Dosi, 1998) into individuals' habituated action to a point beyond rational decision making or deliberate choice (Cohen and Squire, 1980, Winter, 2006).

Strong and coherent routines allow actors to economize on finding facts, processing information and getting things done by simplifying the everyday tasks of making decisions (Heiner, 1983; Penrose, 1959; Simon, 1982). Even intentional knowledge creation by intelligent, self-interested individuals becomes path-dependent as today's routines are related to yesterday's learning routines and knowledge (Arthur, 1994; Hayek, 1960).

Such evolutionary processes of social or technical innovation, selection, and retention[2] often produce what Spender (1996) has labeled 'automatic knowledge'. Those who possess it know little or nothing of its origin or how they have come to know it, but they know how to use it and they can see that it works.

Experience alone does not guarantee any local or general increase in the stock of knowledge, however. On the contrary, stagnation may be a stable condition and the 'solutions to some surprisingly simple technical problems appear to have eluded producers despite centuries of repetitive activity' (Young, 1993, p. 444). Wrong conclusions may be drawn by misinterpretation of the facts at hand (Loasby, 1991), and can even become widely dispersed.

Even if individuals and firms get things right in the first place, they do not always stay that way. Circumstances change and routines that work well at one point in time may later become outdated. Routines tend to be durable, however. They are often retained and defended by actors with vested interests long after changes in the external conditions have made them redundant (Demsetz, 1988). It is thus difficult to unlearn successful habits of the past, even in cases where it is obvious to everyone concerned that they hinder future success (Hedberg, 1981). Firms may find themselves in lock-in situations for long periods, until some entrepreneurial actor, inside or outside the firms in question, unrestrained by the dominant disposition, breaks the spell by introducing new ways of doing things.

Research in cognitive psychology addresses the behavioral drivers and constraints at the micro level, especially concerning human judgment and decision-making under uncertainty. There seems to be a general set of biases in individuals' search among the wide range of opportunities in their external environment (Simon, 1947 and Tversky and Kahneman, 1974). Bounded rationality, sometimes reinforced by attributes of problem architectures and misaligned incentives, makes individuals concentrate their search on a restricted range of potential alternatives (March, 1991; Ocasio, 1997). Local, exploitive search, that is, looking for answers close to already existing solutions while utilizing existing routines, is preferred even in situations when the possible higher search costs of following many different paths simultaneously or pursuing a more global search strategy is more than balanced by the potential benefits of acquiring a broad variety of knowledge inputs (Jensen and Meckling, 1976; Simon, 1987; Tversky, 1972).

Elsewhere, we have used the term 'functional myopia' to denote this form of restricted search for knowledge and solutions (Maskell and Malmberg, 2007) and noted that it has an interesting corresponding spatial aspect (Levinthal and March, 1993). The concept of localized learning (Malmberg and Maskell, 2006; Maskell and Malmberg, 1999) has been coined to demonstrate how some processes of interactive knowledge creation, acquisition and exchange tend to be if not exclusively so at least in some central respects local in character. Despite modern developments in transport and communications technology that make long-distance interaction easier and cheaper, certain types of information and knowledge exchange continue to require regular and direct face-to-face contact (Storper and Venables, 2004; and see Goffman, 1967). Everything else being equal, knowledge search and exchange processes will usually be less costly and smoother, the shorter the physical distance between the participants. Also social affinity, cultural community and cognitive proximity tend to follow from spatial proximity (Gertler, 1995). In this sense, processes of localized learning are inherent in the everyday life of people working – and living – in any local setting.

There are at least three sources of localized learning processes (see Malmberg and Maskell, 1997, 2002). All three relate to time-geography and the fact that spatial proximity makes face-to-face interaction easier (less costly, time consuming, tiresome) and tends to carry with it an element of social, cultural, and not least cognitive, proximity.

The first two sets of sources of localized learning relate to the two dimensions – vertical and horizontal – along which the division of labor may develop. The third is based on the assertion that spatial proximity increases the likelihood of fruitful unanticipated encounters among key players that help them incorporate insights, opinions, and ideas from a broader community of informed observers. This third – social – dimension of spatial proximity may be seen as a neighborhood effect.

In the *vertical dimension*, firms specializing in different stages of a production chain are linked through input/output relations. They possess knowledge, experience, or skills useful for undertaking dissimilar but complementary activities. However, while vertical links make up the core of much analysis of localized learning, the fact remains that few empirical studies have been able to show that vertical linkages are indeed predominantly local (Gordon and McCann, 2005; Malmberg and Maskell, 2002). There are presumably several reasons for this, one being the spatial extension of markets that is part and parcel of the general process of globalization. Successful firms do tend to 'grow out of their local context' and access more distant markets as they develop and grow (this is indeed part of the definition of competitive success). Therefore it should be no surprise that most firms, also if located in vibrant clusters, have most of their suppliers, customers or collaborators in distant markets.

Many localized industrial systems are instead based on *the horizontal dimension*. Such spatial agglomerations are made up of several firms operating in the same industry, producing similar output. While vertically related firms are business partners and collaborators, horizontally related firms are mainly rivals and competitors (Maskell, 2001). Marshall (1890) first drew attention to the possible advantages of variation that are caused by the localized and parallel performance of similar tasks carried out by independent firms. The parallel experimentation and testing of approaches by independent firms – with different perceptive powers, insights, attitudes and assessment of the information at hand – is a constant source of variation, despite the fact that they act within the same institutional structure (see the next section). An essentially Darwinian feature of such variation is that it does not presuppose any trust whatsoever among the firms as a prerequisite for learning. It does not require any close contact or even an arm's-length interaction between the firms, but is simply the result of the parallel experimentation that goes on in a horizontally disintegrated system.

Furthermore, collocated firms undertaking similar activities find themselves in a situation where every difference in the solutions chosen, however small, can be observed and compared. First, regarding *observability*, spatial proximity brings with it the special feature of spontaneous, automatic observation. Business firms often have remarkably good knowledge of the undertakings of nearby firms. A second element is *comparability*. While it might be easy for firms to blame an inadequate local factor market when confronted with the superior performance of competitors located far away, it is less so when the premium producer is located in the same local environment. The sharing of common conditions, opportunities, and threats makes the strengths and weaknesses of each individual firm apparent to anyone who cares to take an interest. Firms with similar capabilities in the horizontal dimension constantly imitate the proven or foreseeable success of one another.

While suppliers and customers in a vertically organized production chain need to interact with each other in order to do business, competitors do not. Introducing the horizontal dimension means that localized learning in part may be independent of the degree of internal interaction, at least in principle. The only requirement is that several firms undertaking similar activities are placed in circumstances where they can monitor and compare each other's undertakings constantly, closely, and almost without effort or costs.

Local competition between similar firms may of course also make up a selection mechanism in its own right. The fact that the co-presence of many similar firms creates a local environment in which it is hard to survive (recent studies show that spatial clustering may bring about a high degree of exits, see e.g. Boschma and Wenting, 2007) can be seen as a cluster disadvantage (as a form of diseconomies of agglomeration) but also as something that fosters globally competitive firms.

In addition to the proximity effects that manifest themselves in interaction or encounters between collocated firms, there are processes of localized learning that are inherent in the everyday life of people working – and living – in any local setting. Information exchange – or knowledge spillovers – does go on that is not related to the conscious undertakings of firms, but is rather to be seen as an unintended side effect of such undertakings. By taking up a concept introduced by Storper and Venables (2004), the authors have previously denoted this phenomenon as 'local buzz' (Bathelt et al., 2004). In a similar way, Owen-Smith and Powell (2004) use the notion of 'local broadcasting' and Grabher (2002) the term 'noise' to denote the idea that there are lots of useful things going on simultaneously in a local milieu and therefore lots of inspiration and information to receive for the perceptive local actor. Buzz refers to the information and communication ecology created by numerous face-to-face contacts as people and firms within the same industry collocate in the same city, district, or region. Buzz consists of specific information and continuous updates of this information; intended and unanticipated learning processes in organized and accidental meetings; the application of the same interpretative schemes and mutual understanding of new knowledge and technologies; as well as shared cultural traditions and habits, which taken together make interaction and learning less costly. Actors continuously contribute to and benefit from the diffusion of information, gossip, and news by just dwelling in a certain place (Gertler, 2003).

Firms that are located in the same place are also able to understand the local buzz in a meaningful and useful way. The reason is, as already noted, that spatial proximity stimulates firms to develop similar language, technology attitudes, and interpretative schemes (Lawson and Lorenz, 1999; Maskell and Lorenzen, 2004). Also, as has been suggested elsewhere (Maskell et al., 1998), trust tends to exist in local milieus as something inherited, that any 'insider' will benefit from by default. Local milieus thus consist of agents that are bound together by day-to-day interaction, based on the same expertise, a common set of technological knowledge, and similar experience with a particular set of problem-solving techniques. Such relations can develop within a firm but also span a single organization and include other firms of a value chain (Gertler, 2001).

These processes can be seen as analogous with the neighborhood effect that has been established, for example, in studies of voting behavior. In that context Johnson et al.

(2005) recently identified five separate processes that produce the same outcome (i.e. they contribute to making people in a local area behave in a certain – uniform or at least similar – way). In addition to the classic neighborhood effect based on local social interaction, there may be process of emulation, environmental observation, environmental selection and local pressure, respectively (see Malmberg and Maskell, 2006).

All in all, there seems to be reasonable support for the claim that spatial proximity does make possible certain types of interactive learning. A division of processes of localized learning into the vertical, the horizontal, and the social realm is a useful way to identify in more detail how they come about, and they all point in the direction of supporting the notion of myopia in processes of knowledge search. Vicarious localized learning processes allegedly lead to spatial myopia in the sense that they contribute to direct search processes into local, isomorphic paths (Levitt and March, 1988).

In section 4 we return to the issue of whether certain types of local milieu, notably clusters, have characteristics that may, at least in part, help compensate for the negative aspects of myopia and perhaps even turn them into advantages instead. But before doing so, we direct our attention to the macro-level phenomena of institutions and institutional change, and how they contribute to path-dependent development at the aggregate level of cities, regions and nations.

3. Higher-order processes affecting knowledge creation: the role of institutional settings

Institutions create incentives and guidelines for action (Scott, 2003). They can be conceived of as resilient humanly devised constraints that structure interaction in society (North, 1994). These constraints may be regulative (e.g. formal rules, laws, constitutions), normative (e.g. norms of behavior, conventions, self-imposed codes of conduct), or cultural-cognitive (e.g. creating interpretive frameworks by which meaning is established and shaped). It is commonly recognized that institutions do not come into existence once and for all. On the contrary they undergo continual modifications to adjust to the circumstances of time and place, so that the restraints imposed on individual action change over time.

Institutions can be thought of at different levels of aggregation. At the most aggregate level, the common set of market economy institutions provides individuals with incentives to experiment with 'new combinations' (Schumpeter, 1934, p. 66) of existing knowledge, while occasionally adding a novel idea or two. In other words: they jointly create an encouraging environment for the whole evolutionary sequence of innovation, selection and retention.

The perhaps most significant consequence of market economy institutions is the way they allow for divergent interpretations of potentially successful behavior and the continued coexistence of dissimilar approaches at the micro level (Marshall, 1890). The resulting variation in individual responses to shared challenges and opportunities in the environments provides material for the subsequent process of selection and interactive knowledge creation.

At the less aggregate level of nations, regions or cities, variation may appear in the general institutions of the experimenting market economy. Initially random and insignificant territorial nuances in institutional repertoires deepen over time in response to the requirements of existing economic configurations. They thereby, in turn, create a particularly favorable environment for attracting newcomers engaged in precisely those specific kinds of activity (Maskell et al., 1998). In many small steps, distinctive and resilient institutional combinations are being established along national, regional or local lines (Hall and Soskice, 2001; Lundvall and Maskell, 2000; Whitley, 1995, 1999). The national, regional and local scale will typically display only partially overlapping institutional repertoires, but the general process of alignment over time between institutional set-up and economic structure is equally important at each scale (Gertler, 2001).

An interesting aspect of institutions in general is that they tend to work by limiting or even preventing the exploration of alternative possibilities (Loasby, 2000). A historically evolved institutional repertoire can thus sometimes lead whole countries, regions or cities into specific, initially successful, ways of doing things that later external events convert into shackles that inhibit or block further progress (Elbaum and Lazonick, 1986). At this aggregate level competition is not to be relied on to ensure rejuvenation. Some nations, regions or cities instead depopulate or accept, however reluctantly, a continuous decline in investment levels, consumption and standards of living. Others, however, acknowledge that some collective action is required and struggle to develop deliberate policies often aimed at emulating the institutional structure of more successful peers (Czarniawska and Joerges, 1996). Many studies have shown how new ideas, organizational forms or legislative practices travel across space while undergoing modifications or 'translations' to fit the new setting (Czarniawska and Joerges, 1996; Scott, 2003). Investments in increased mobility or extending relational systems might make ideas travel more readily, but the crucial point is the extent to which such alien ideas can be absorbed into the national, regional or local institutional setup and, by becoming part thereof, revitalize it and help break the lock-in.

It should be admitted straight away that we know only a little about what determines institutional absorptive capacity and why the institutional set-up of nations, regions or cities on the downhill slope sometimes remains unaffected through extended periods in spite of seemingly infinite opportunities for external inspiration from more successful peers. In particular, the literature offers limited insight into precisely *how* to import unfamiliar institutions that are deemed to be superior.

There are probably several reasons for this lacuna, but the most obvious is, perhaps, an inherent logical problem. Ingrained institutions, by definition, function behind the back, so to speak, of the individual actor. They confine his or her actions by guiding their mindset into certain ways of thinking. But if an institution makes certain actions seem natural and self-evident for the individuals constrained by them, then how can these same people, by an act of will, ever step outside an innate institution and change it? Only recently have scholars in social science started to come to grips with the fundamental question of how obsolete but well-established institutions can intentionally be transformed at the micro-level of individual action (see Schneiberg, 2007 for a review of recent advances in this area).

A similar theoretical gap concerns the link connecting institutional change to broader external events like major shifts in technologies or consumer needs. We are still mainly in the dark when attempting to explain why cities, regions or nations sometimes successfully manage to reinvent vital parts of their institutional structure to accommodate external shocks, while sometimes they do not. Related to this is the issue of whether radically new economic activities need radically new institutions, that emerge alongside or instead of old institutions, while more incremental change in the economic activity will merely need the adjustment of old institutions (see the Schumpeterian question of whether major innovations need new or old firms to develop).

It is somewhat discomforting that the same factors – relating to knowledge, routines and institutions – tend to be used to explain both regional success and failure. Until today, economic geographers, as well as their colleagues in the neighboring economic and social sciences, have been only partly successful when it comes to identifying – let alone predicting – the branching points when successful specialization of cities, regions or nations turns into negative lock-in. At one point in time, we may establish as a fact that, for example, a specific form of industrial relations or some other institutional arrangement is major a cause of the industrial growth and dynamism of a city, region or nation. At a later point, we may discover that the very same institutional arrangement creates inertia and prevents the industry of the region from successfully meeting the challenge posed by the emergence of new competitors elsewhere. Such analyses run the risk of getting stuck in a functionalist type of explanation that ends up very close to circular reasoning: the cities, regions or nations that do well are those that are equipped with the appropriate institutional structure, while the definition of an appropriate or even superior institutional structure is the one found in territories that perform well.

The problems of how to build an empirically justified theoretical understanding of institutional dynamics at the level of cities, regions or nations in relation to individual agency on the one hand, and external shocks on the other, remain interesting challenges to evolutionary economic geography.

4. An evolutionary approach to the cluster life cycle

Having identified myopia as a key micro-level constraint in knowledge development, and pinpointed institutional adjustment – or lack of adjustment – as a key factor in directing knowledge processes at the aggregate level of cities, region or nations, in this section we apply these concepts to a paradigmatic economic geographical case at the meso level: that of the rise, growth, decline and possible rejuvenation of spatial clusters of similar and complementary economic activity.

Much may be learnt about the role of space and place in economic processes by trying to pinpoint the driving forces that make for agglomeration in space of economic activity. Indeed, some would claim that spatial clustering is the phenomenon that ultimately motivates and defines the existence of economic geography as an academic discipline (Fujita et al., 1999, p. 4).

Early attempts to define the economic advantages gained by agglomerative behavior were framed by a division into two sorts of agglomeration economies (see Dicken and Lloyd, 1990; Hoover, 1937). Thus, urbanization economies came to denote those advantages to be gained by location in a large and dense urban area, while the notion of localization economies referred to advantages gained by locating close to other similar or related firms.

It is the latter aspect of agglomeration that makes up the main focus of this section. A note on terminology should be entered here. The phenomenon of spatial agglomeration of similar and related economic activity is dealt with in the literature under different headings. Alfred Marshall's (1890) classical – and still in many ways unsurpassed contribution to the topic – uses the concept of localization to analyze the coming into existence of 'industrial districts'. In recent years, clusters and clustering have been the most widely circulated terms, largely following the work by Porter (1990, 1998). It has been argued by some that the agglomeration concept would refer exclusively to the spatial concentration of firms and industries, while the cluster concept would refer to spatial collocation in combination with functional linkage between the parties involved. For our purpose here, however, spatial clustering is the preferred term, defined in the simplest possible way: a cluster is a spatial agglomeration of firms with similar and complementary competencies (Richardson, 1972).

The use of the cluster concept and related models in academic and policy circles has grown immensely over the last decade. The pandemic diffusion of the concept has made it increasingly vague and indeed fuzzy (Malmberg and Maskell, 2002; Malmberg and Power, 2006; Maskell and Kebir, 2006). If there is such an entity as 'spatial clustering/ agglomeration theory' today, it consists of a mix of ideas originating more than a century ago (Marshall, 1890; Weber, 1909), further developed in the post-war period within the regional science tradition (Alonso, 1964) and given new input since the early 1990s though the engagement by economists (notably Krugman, 1991) business strategists (notably Porter, 1990) and economic geographers (see Malmberg and Maskell, 2002 for an overview).

The mechanisms behind spatial clustering identified in traditional cluster theory and the factors in most contemporary models all take their point of departure in the concept of agglomeration, or localization economies (Maskell, 2001). These bodies of thought identify forces that give permanent advantages to firms with similar or complementary competencies when located in a spatial agglomeration. The advantages include, among other things, lower transaction costs, access to a pool of skilled labor, shared costs for specialized infrastructures, and access to knowledge arriving from the outside world through the cluster's global pipelines (Bathelt et al., 2004).

An evolutionary approach will supplement analyses of localization economies by placing the argument about which clusters exist within a sequence that emphasizes how clusters originate, develop and decline. Clusters do indeed often follow an evolutionary path, where stages of infancy are succeeded by a growth phase, followed in turn by increasing maturity and subsequent stages of stagnation or decline. In this sequence they often, but not always, follow the life-cycle of the dominant industry quite closely (Audretsch and Feldman, 1996; Buenstorf and Klepper, 2005). A stylized version of a 'cluster life cycle' can be described as follows (see Press, 2006).

The origin of clusters remains largely obscure, in the sense that it is almost impossible to determine ex ante where a cluster in the making will take root. Sometimes it is obvious that the pre-existence of certain physical factors, such as a localized supply of energy or some raw material, triggered the initial location of an industrial activity in a place that would later on develop into a fully fledged cluster. More often, though, they start out in a particular location more or less by chance and can therefore not be predetermined. It is often possible to trace the roots of a cluster ex post, of course, by means of genealogical analysis. 'It all started' when this or that pioneering entrepreneur at some point in time did start this or that type of economic activity, even though it is not really possible to explain why this happened in one particular place and not in another, apart from the banal fact that most new firms start at the place of birth or residence of the founder.[3] While the incident that actually triggers the clustering process often remains unclear, the subsequent stages are easier to account for, and once an activity has taken root in a particular place, several mechanisms come into play at the micro level of explicit agency.

A potential entrepreneur will decide to start a business similar or complementary to what already exists locally, because (myopic) cognitive limitations reinforce powerful economic incentives. Firms with similar competencies are attracted because collocation implies economizing on locational search costs: the prior local existence of one or more successful firms in an industry proves that no obvious or obscure locational factor makes the area less suitable for that specific kind of economic activity. Collocation furthermore enables latecomers to piggyback on pioneering firms' investments in labor market development, infrastructure adjustments and institution building. Firms with complementary competencies will, in turn, become attracted by the commercial possibilities of increased local demand or the opening of new local sources of supply. A chain of cumulative causation (Myrdal, 1957; Veblen, 1898) can be set in motion where the vertical expansion of the cluster through attraction of firms with complementary competencies adds to the initial attractiveness of the cluster along the horizontal dimension of firms with similar competencies.

The decisions made by the firms have a parallel at the level of individual workers, as the most talented wannabes of an industry will tend to flock around the hotspots where firms in this line of business agglomerate. But over time, some employees will leave the incumbent firms to start their own business while utilizing firm-specific routines learnt before the spin-off. Particularly successful routines would thus not only provide parent firms with above-average chances of survival but, consequentially, also enable numerous spin-offs, each endowed with inherited and previously tried-out routines. Through rounds of selection and intensified local rivalry (Porter, 1990) the spin-offs help create stable clusters with firms that through generations can dominate an industry nationally or even globally (Boschma and Wenting, 2007; Buenstorf and Klepper 2005; Thompson and Klepper, 2005; Klepper, 2002; Dahl et al., 2003, 2005).

The assertion that clusters are essentially the result of myopic behavior should thus not be taken to indicate that clusters cannot for shorter or longer periods be efficient, or indeed competitive, spatio-organizational forms. On the contrary, clusters represent a possibility to circumscribe some of the potential problems stemming from myopia. The variety that follows from the parallel experimentation going on when a number of firms set out to solve similar problems while competing locally can be seen as a localized capability. If functional and spatial myopia is indeed a human condition, those actors lucky enough to find world-class solutions and global best practices in their backyard gain advantages over those who do not. In this way, a set of decisions and actions, partly framed by the cognitive and other constraints of the actors involved, can combine to create an aggregate structure that turns out to be not just economically sustainable but indeed globally successful.

But, as shown in previous sections, the success also carries with it the seeds of future destruction as the evolutionary process of selecting temporarily best practices accumulate to isomorphic pressures that gradually reduce existing variety in routines. Lack of variety combined with spatial myopia leads to an insular mind-set that, in turn, enables local actors to ignore signs of needed readjustment. However, the potentially most damaging long-term consequences may be avoided as long as at least some of the collocated firms actively invest in building absorptive capacities and pipelines to external knowledge pools with dissimilar routines or institutional patterns. Pending lock-ins are replaced by rejuvenation processes through the activities of externally connected and absorptive local firms that ensure variety and create fresh impulses for horizontal learning.

In addition to micro-level processes, a cluster's life cycle account would also include higher-order processes at the level of the cluster taken as a whole. These include, for example, the creation of supportive institutions, local culture and the establishment of the place as a brand of the dominant local industries. Such factors reinforce the specialization of the cluster, and by doing so they set the frame for the kind of activities that might be possible to perform in the future in that area.

The evolutionary chain of cumulative causation favors industrial specialization and territorial differentiation in dominant routines and institutional repertoires that help explain why no competitive cluster – or city or region or nation for that matter – can remain a Jack-of-all-trades. At the same time, both specialization and differentiation come at a cost. On the one hand there are beneficial effects of deepening the knowledge base and developing routines that work well within a particular field. On the other hand clusters run the risk of getting locked in if external events render the knowledge base obsolete, the dominant routines dysfunctional, or the institutional set-up inflexible (Cornwall and Cornwall, 2001).

In essence, an evolutionary approach directs our attention to the mechanisms by which a pattern of clusters would emerge through the actions of myopic prospective entrepreneurs, grow through the gradual build-up of well-adjusted routines and institutions, and face threats of decline as a result of over-specialization, leading to reduced variety, a narrowing down of knowledge bases and the rise of inertia. The standard accounts of permanent localization economies – the cluster existence arguments – are in this way placed in an evolutionary sequence (see Maskell and Kebir, 2006).

5. Evolutionary economic geography: conclusions and challenges

The argument forwarded in this chapter can be summarized in four bold points.

Our main argument is that the explicit recognition of the specific drivers and constraints guiding individual action does improve our understanding of aggregate processes also at the levels of cities, regions or nations. Thereby, we have attempted to demonstrate how an evolutionary approach has much to offer, but have also pointed to a number of unresolved issues or major theoretical challenges for future theory building. Whether the promises raised by the still juvenile evolutionary economic geography will be fulfilled as it matures depends on its ability to invigorate empirical research on economic development of nations, regions, cities and, indeed, of clusters – not least when it comes to explaining how and why institutional change comes about.

Second, we have demonstrated how the adoption of an evolutionary approach might enhance our understanding of the birth, growth and decline of clusters. Local entrepreneurs will tend to make choices that reinforce a pattern of specialization, whatever the precise cause of the initial spark happened to be. Institutional adjustment gradually increases the fit with the chosen specialization and adds to the performance of the cluster. It is an unavoidable consequence of this process that it reduces the range of alternative development paths that may become attractive when external circumstances take unexpected turns. The immediate benefits of deepened specialization are balanced by increase in the risk of lock-in, stagnation or even decline.

Third, we have attempted to show how evolutionary processes of social or technical innovation, selection, and retention lead to the gradual build-up of routines that allow actors to economize on fact-finding and information processing. A strong and coherent array of routines simplifies the everyday tasks of making decisions. At the same time, there are cognitive constraints at the micro level. Limited cognitive abilities make individuals prefer local, exploitive search, in two different ways. They tend to look for solutions close to already existing routines, but they also tend to concentrate their search in their spatial vicinity. The incorporation of functional and/or spatial myopia as a basic behavioral assumption implies departing from mainstream economic conjectures of rationalization, global maximization and equilibria.

Finally, leaving the micro-level phenomena and turning to the level of markets and institutional regimes, we have maintained that a dominant institutional set-up, once established, would tend to attract those firms and individuals most compatible with it. National, regional or local institutions gradually develop over time in response to the special requirements of the presently dominating industry and lead to further specialization by creating a favorable environment for similar and complementary economic activity. Thus, at the aggregate level of cities, regions or nations, evolutionary economic geography can be argued to have as its main object of study the inertia and path dependence caused by the development, reproduction and transformation of territorially specific routines and institutions.

Acknowledgement

This chapter partly draws on previous work, notably Maskell and Malmberg (1999), Malmberg and Maskell (2002) and Maskell and Malmberg (2007). The first few paragraphs of this chapter also appear, in slightly modified form, in Lagerholm and Malmberg (2009). We acknowledge supportive as well critical comments from the editors and anonymous referees in the process of preparing and revising this chapter.

Notes

1. While focusing on knowledge evolution, this chapter does not intend to give a full account of the large and growing literature on the general economics of knowledge, following in the wake the founding papers by Nelson (1959) and Arrow (1962). Basic notions like knowledge appropriability, indivisibility or non-exhaustability are not explicitly considered in any detail here. These issues, as well as the distinction between tacit and codified knowledge, have been addressed in other papers by the authors, on which this chapter builds (see e.g. Malmberg and Maskell, 2006; and Maskell and Malmberg, 1999, 2007). For a broad overview, see also Gertler (2003).
2. The sequence is a common social science substitute for the initial Darwinian terms of mutation, selection and transmission; see Essletzbichler and Rigby (2007) or Glückler (2007).
3. Krugman's first engagement with economic geography starts out with a famous example of this (Krugman, 1990).

References

Agassi, J. (1960), 'Methodological individualism', *British Journal of Sociology*, **11**, 244–70.

Alonso, W. (1964), *Location and Land Use: Toward a General Theory of Land Rent*, Cambridge, MA: Harvard University Press.

Arrow, K.J. (1962), 'The economic implications of learning by doing', *Review of Economic Studies*, **29**, 155–73.

Arthur, W.B. (1989), 'Competing technologies, increasing returns, and lock-ins by historical events', *Economic Journal*, **99**, 116–31.

Arthur, W.B. (1994), *Increasing Returns and Path Dependence in the Economy*, Ann Arbor: The University of Michigan Press.

Audretsch, D.B. and M.P. Feldman (1996), 'Innovative clusters and the industry life cycle', *Review of Industrial Organization*, **11**, 253–73.

Bathelt, H., A. Malmberg and P. Maskell (2004), 'Clusters and knowledge: local buzz, global pipelines and the process of knowledge creation', *Progress in Human Geography*, **28**, 31–56.

Boldrin, M. and J.A. Scheinkman (1988), 'Learning-by-doing, international trade and growth: a note', in P.W. Anderson, K.J. Arrow and D. Pines (eds), *The Economy as an Evolving Complex System*, Reading, MA: Addison-Wesley, The Advanced Book Program, pp. 285–300.

Boschma, R. and K. Frenken (2006), 'Why is economic geography not an evolutionary science? Towards an evolutionary economic geography', *Journal of Economic Geography*, **6**, 273–302.

Boschma, R.A. and R. Wenting (2007), 'The spatial evolution of the British automobile industry. Does location matter?', *Industrial and Corporate Change*, **16** (2), 213–38.

Buenstorf, G. and S. Klepper (2005), 'Heritage and agglomeration: the Akron tire cluster revisited', Papers on Economics and Evolution No. 2005–08, Max Planck Institute of Economics, Evolutionary Economics Group.

Cohen, N.J. and L.R. Squire (1980), 'Preserved learning and retention of pattern-analyzing skill in amnesia: dissociation of knowing how and knowing that', *Science*, **210**, 207–10.

Coleman, J.S. (1986), 'Social theory, social research, and a theory of action', *American Journal of Sociology*, **91**, 1309–35.

Coleman, J.S. (1990), *Foundations of Social Theory*, Cambridge, MA: Belknap Press.

Coriat, B. and G. Dosi (1998), 'The institutional embeddedness of economic change: an appraisal of the "evolutionary" and "regulationist" research programmes', in K. Nielsen and B. Johnson (eds), *Institutions and Economic Change – New Perspectives on Markets, Firms and Technology*, Cheltenham, UK and Lyme, NH, USA: Edward Elgar, pp. 3–32.

Cornwall, J. and W. Cornwall (2001), *Capitalist Development in the Twentieth Century: An Evolutionary Keynesian Analysis*, Cambridge: Cambridge University Press.

Czarniawska, B. and B. Joerges (1996), 'Travels of ideas', in B. Czarniawska and G. Sevón (eds), *Translating Organizational Change*, Berlin: Walter de Grüyter, pp. 13–48.

Dahl M.S., C.Ø.R. Pedersen and B. Dalum (2003), 'Entry by spinoff in a high-tech cluster', DRUID Working Paper No. 2003–11.

Dahl, M.S., C.Ø.R. Pedersen and B. Dalum (2005), 'Entrepreneurial founder effects in the growth of regional clusters: how early success is a key determinant', DRUID Working Paper No. 2005–18.

Demsetz, H. (1988), 'The theory of the firm revisited', *Journal of Law, Economics and Organization*, **4**, 141–61.

Dicken, P. and P.E. Lloyd (1990), *Location in Space: Theoretical Perspectives in Economic Geography*, 3rd edn, New York: Harper & Row.

Elbaum, B. and W. Lazonick (eds) (1986), *The Decline of the British Economy*, Oxford: Clarendon Press.

Essletzbichler, J. and D.L. Rigby (2007), 'Exploring evolutionary economic geography', *Journal of Economic Geography*, **7**, 549–71.

Frenken, K. and R. Boschma (2007), 'A theoretical framework for evolutionary economic geography. Industrial dynamics and urban growth as a branching process', *Journal of Economic Geography*, **7**, 635–49.

Fujita, M., P. Krugman and A.J. Venables (1999), *The Spatial Economy*. *Cities, Regions and International Trade*, Cambridge, Ma & London: MIT Press.

Gertler, M.S. (1995), 'Being there – proximity, organization, and culture in the development and adoption of advanced manufacturing technologies', *Economic Geography*, **71**, 1–26.

Gertler, M.S. (2001), 'Best practice? Geography, learning and the institutional limits to strong convergence', *Journal of Economic Geography*, **1**, 5–26.

Gertler, M.S. (2003), 'Tacit knowledge and the economic geography of context, or the undefinable tacitness of being (there)', *Journal of Economic Geography*, **3**, 75–99.

Glückler, J. (2007), 'Economic geography and the evolution of networks', *Journal of Economic Geography*, **7**, 619–34.

Goffman, E. (1967), *Interaction Rituals: Essays on Face-to-Face Behavior*, Garden City, NY: Doubleday & Co.

Gordon, I.R. and P. McCann (2005), 'Innovation, agglomeration, and regional development', *Journal of Economic Geography*, **5** (5), 523–53.

Grabher, G. (2002), 'The project ecology of advertising: tasks, talents and teams', *Regional Studies*, **36** (3), 245–62.

Hall, P.A. and D. Soskice (2001), *Varieties of Capitalism: The Institutional Foundations of Comparative Advantages*, Oxford and New York: Oxford University Press.

Hayek, F.A. von (1960), *The Constitution of Liberty*, Chicago: University of Chicago Press.

Hedberg, B. (1981), 'How organizations learn and unlearn', in P.C. Nystrøm and W.H. Starbuck (eds),

Handbook on Organizational Design – Adapting Organizations to their Environment, Oxford: Oxford University Press, pp. 3–27.
Heiner, R.A. (1983), 'The origin of predictable behavior', *American Economic Review*, **73**, 560–95.
Hirschman, A.O. (1958), *The Strategy of Economy Development*, New Haven, CT: Yale University Press.
Hoover, E.M. (1937), *The Location Theory and the Shoe and Leather Industries*, Cambridge, MA: Harvard University Press.
Jensen, M.C. and W. Meckling (1976), 'The theory of the firm: managerial behavior, agency costs and ownership structure', *Journal of Financial Economics*, **3**, 305–60.
Johnston, R., C. Propper, R. Sarker, K. Jones, A. Bolster and S. Burgess (2005), 'Neighbourhood social capital and neighbourhood effects', *Environment and Planning A*, **37** (8), 1443–59.
Klepper, S. (2002), 'The capabilities of new firms and the evolution of the US automobile industry', *Industrial and Corporate Change*, **11**, 645–66.
Krugman, P.R. (1990), *Geography and Trade*, Leuven, Belgium & Cambridge, MA: Leuven University Press and MIT Press.
Krugman, P.R. (1991), 'Increasing returns and economic geography', *Journal of Political Economy*, **99**, 483–99.
Lagerholm, M. and A. Malmberg (2009), 'Path dependence in economic geography', in L. Magnusson and J. Ottosson (eds), *The Evolution of Path Dependence*, Cheltenham, UK and Northampton, MA, USA: Edward Elgar.
Lawson, C. and E. Lorenz (1999), 'Collective learning, tacit knowledge and regional innovative capacity', *Regional Studies*, **33** (4), 305–17.
Levinthal, D.A. and J.G. March (1993), 'The myopia of learning', *Strategic Management Journal*, **14**, 95–112.
Levitt, B. and J.G. March (1988), 'Organizational learning', *Annual Review of Sociology*, **14**, 319–40.
Liebowitz, S.J. and S.E. Margolis (1995), 'Path dependence, lock-in and history', *Journal of Law, Economics and Organization* (April), 205–26.
Loasby, B.J. (1991), *Equilibrium and Evolution. An Exploration of Connecting Principles in Economics*, Manchester: Manchester University Press.
Loasby, B.J. (2000), 'Market institutions and economic evolution', *Journal of Evolutionary Economics*, **10**, 297–309.
Lundvall, B.-Å. and P. Maskell (2000), 'Nation states and economic development – from national systems of production to national systems of knowledge creation and learning', in G.L. Clark, M.P. Feldman and M.S. Gertler (eds), *The Oxford Handbook of Economic Geography*, Oxford: Oxford University Press, pp. 353–72.
Malmberg, A. and P. Maskell (1997), 'Towards an explanation of regional specialization and industry agglomeration', *European Planning Studies*, **5**, 25–41.
Malmberg, A. and P. Maskell (2002), 'The elusive concept of localization economies: towards a knowledge-based theory of spatial clustering', *Environment and Planning A*, **34**, 429–49.
Malmberg, A. and P. Maskell (2006), 'Localized learning revisited', *Growth and Change*, **37**, 1–18.
Malmberg, A. and D. Power (2006), 'True clusters. A severe case of conceptual headache', in B.T. Asheim, P. Cooke and R.L. Martin (eds), *Clusters and Regional Development*, London: Routledge.
March, J.G. (1991), 'Exploration and exploitation in organization learning', *Organization Science*, **2**, 71–87.
Marshall, A. (1890), *Principles of Economics*, London: Macmillan.
Martin, R. and P. Sunley (2006), 'Path dependence and regional economic evolution', *Journal of Economic Geography*, **6**, 395–437.
Maskell, P. (2001), 'Towards a knowledge-based theory of the geographical cluster', *Industrial and Corporate Change*, **10**, 919–41.
Maskell, P., H. Eskelinen, I. Hannibalsson, A. Malmberg and E. Vatne (1998), *Competitiveness, Localised Learning and Regional Development. Specialisation and Prosperity in Small Open Economies*, London: Routledge.
Maskell, P. and L. Kebir (2006), 'The theory of the cluster – what it takes and what it implies', in B.T. Asheim, P. Cooke and R.L. Martin (eds), *Cluster and Regional Development: Critical Reflections and Explorations*, London: Routledge, pp. 30–49.
Maskell, P. and M. Lorenzen (2004), 'The cluster as market organization', *Urban Studies*, **41** (5–6), 991–1009.
Maskell, P. and A. Malmberg (1999), 'Localised learning and industrial competitiveness', *Cambridge Journal of Economics*, **23** (2), 167–185.
Maskell, P. and A. Malmberg (2007), 'Myopia, knowledge development and cluster evolution', *Journal of Economic Geography*, **7**, 603–18.
Myrdal, G. (1957), *Economic Theory and Under-Developed Regions*, London: Duckworth.
Nelson, R.R. (1959), 'The simple economics of basic scientific research', *The Journal of Political Economy*, **67** (3), 297–306.
North, D.C. (1994), 'Institutional competition', in H. Siebert (ed.), *Locational Competition in the World Economy*, Tübingen: J.C.B. Mohr (Paul Siebeck), pp. 27–44.

Ocasio, W. (1997), 'Toward an attention-based view of the firm', *Strategic Management Journal* (Special Issue), **18**, 187–206.

Owen-Smith, J. and W.W. Powell (2004), 'Knowledge networks as channels and conduits: the effects of formal structure in the Boston biotechnology community', *Organization Science*, **15** (1), 5–21.

Penrose, E.T. (1959), *The Theory of the Growth of Firms*, New York: Oxford University Press.

Porter, M.E. (1990), *The Competitive Advantages of Nations*, London/Basingstoke: Macmillan.

Porter, M.E. (1998), 'Clusters and the new economics of competition', *Harvard Business Review*, **76**, 77–90.

Press, K. (2006), *A Life Cycle for Clusters? The Dynamics of Agglomeration, Change and Adaptation*, Heidelberg: Physica-Verlag.

Richardson, G.B. (1972), 'The organisation of industry', *Economic Journal*, **82**, 883–96.

Schneiberg, M. (2007), 'What's on the path? Path dependence, organizational diversity and the problem of institutional change in the US economy, 1900–1950', *Socio-Economic Review*, **5** (2), 47–80.

Schumpeter, J.A. (1934), *The Theory of Economic Development. An Inquiry into Profits, Capital, Credit, Interest and the Business Cycle*, Cambridge, MA, Harvard University Press. (Translated from the original work *Theorie der wirtschaftlichen Entwicklung*, München, 1911).

Scott, W.R. (2003), 'Institutional carriers: reviewing modes of transporting ideas over time and space and considering their consequences', *Industrial and Corporate Change*, **12**, 879–94.

Scribner, S. (1986), 'Thinking in action: some characteristics of practical thought', in R.J. Sternberg and R.K. Wagner (eds), *Practical Intelligence. Nature and the Origins of Competence in the Everyday World*, Cambridge: Cambridge University Press, pp. 13–30.

Simon, H.A. (1947), *Administrative Behavior: A Study of Decision-Making Process in Administrative Organization*, Chicago: Macmillan.

Simon, H.A. (1982), *Models of Bounded Rationality I: Economic Analysis and Public Policy*, Cambridge MA: MIT Press.

Simon, H.A. (1987), 'Behavioral economics', in J. Eatwell, M. Millgate and P. Newman (eds), *The New Palgrave: A Dictionary of Economics*, London: Macmillan, pp. 221–5.

Spender, J.-C. (1996), 'Competitive advantage from tacit knowledge? Unpacking the concept and its strategic implications', in B. Moingeon and A. Edmondson (eds), *Organizational Learning and Competitive Advantage*, London: Sage.

Storper, M. and A.J. Venables (2004), 'Buzz: face-to-face contact and the urban economy', *Journal of Economic Geography*, **4**, 351–70.

Thompson, P. and S. Klepper (2005), 'Spinoff entry in high-tech industries: motives and consequences', Working Paper No. 0503, Florida International University, Department of Economics.

Thorndike, E.L. and R.T. Rock (1934), 'Learning without awareness of what is being learned or intent to learn it', *Journal of Experimental Psychology*, **17**, 1–19.

Tversky, A. (1972), 'Elimination by aspects: a theory of choice', *Psychological Review*, **79**, 281–99.

Tversky, A. and D. Kahneman (1974), 'Judgment under uncertainty: heuristics and biases', *Science*, **185**, 1124–31.

Ullman, E.L. (1958), 'Regional development and the geography of concentration', *Papers and Proceedings of the Regional Science Association*, **IV**, 179–98.

Veblen, T. (1898), 'Why is economics not an evolutionary science?', *Quarterly Journal of Economics*, **12**, 373–97.

Weber, A. (1909), *Über den standort der industrien*, Translated by Carl Joachim Freidrich and published 1929 under the title: *Theory of the Location of Industries*, Chicago: University of Chicago Press, Tübingen: J.C.B. Mohr.

Whitley, R. (ed.), (1995), *European Business Systems. Firms and Markets in their National Contexts*, London: Sage Publications.

Whitley, R. (1999), *Divergent Capitalisms. The Social Structuring and Change of Business Systems*, Oxford: Oxford University Press.

Winter, S.G. (2006), 'Toward a neo-Schumpeterian theory of the firm', *Industrial and Corporate Change*, **15**, 125–41.

Young, A. (1993), 'Invention and bounded learning by doing', *Journal of Political Economy*, **101**, 443–71.

19 Path dependence and path plasticity: the co-evolution of institutions and innovation – the German customized business software industry

Simone Strambach

1. Introduction

Path dependence and the co-evolution of technology and institutions are key concepts in understanding the dynamics of structural change at the level of firms, sectors and multilevel spatial scales. On the research agenda of evolutionary economic geography, path dependence appears as one of the core components, although there are unresolved issues by its application in economic geography (Martin and Sunley, 2006, and Chapters 3 and 4 in this book; see also Boschma and Martin, 2007, and Boschma and Frenken, Chapter 5 in this book).

Institutional change and institutional dynamics within path-dependent developments is one of the under-explored areas. The chapter argues that for an evolutionary approach in economic geography there is a need to recognize and conceptualize institutional arrangements and institutional change in greater depth. Especially for the understanding of long-term dynamics of economies in space and time, it is necessary to make them an integral part of the analysis and to explain how institutions play a role in dynamic developments of evolutionary paths.

The concept of path dependency is used in economic geography to explain the economic specialization and longstanding success, as well as the crises and economically unfavourable development of regions. The contribution of institutions, 'institutional thickness' and the place-specific institutional settings are made responsible for positive lock-in effects and the ability of regions to adapt continually to a changing economic environment. On the other hand, the place-specific institutional endowment and institutional inertia are also utilized to explain why some regions are victims of their past economic success or cannot escape previous lock-ins (Grabher, 1993; Hassink, Chapter 21 in this book). The understanding of the institutional dynamics within a well-established technological and institutional development path of territorial settings is a central but to a large extent also open issue (see also Maskell and Malmberg, 2007, and Malmberg and Maskell, Chapter 18 in this book).

In recent years, the concept of 'path creation' assumes that paths can be deliberately created by actors if they are able to mobilize the necessary resources for the breakthrough (Garud and Karnøe, 2001). Taking into account that modes of institutional change underpin path dependency and path creation, both approaches have different perspectives. The concept of path dependency emphasizes the institutional functions supporting continuity by stabilizing behaviour and guiding actions of actors, while the concept of path creation focuses on the 'creative destruction' and underlines the breaking of institutional stability and the creation of new institutions for further innovation.

The chapter sheds light on a different mode of institutional change using the notion of 'path plasticity'. Path plasticity does not contradict path dependency or the option of deliberate path creation, but argues that paths are not coherent in themselves. There is 'path plasticity', which describes a broad range of possibilities for the creation of innovation within a dominant path of innovation systems (Strambach and Storz, 2008). Plasticity results, among other things, from the elastic stretch of institutions and institutional arrangements and their interpretative flexibility through actors. It is not always necessary to break through path dependency or to 'lock out': instead the chapter argues that even innovation with a minor degree of complementarity within the well-established institutional setting of technological development paths may come into being through interpretative flexibility and the crawling nature of institutions that results in incremental change. Geography does play an important role in processes of exploring and exploiting plasticity within a well-established institutional setting through place-specific characteristics, processes of localized learning, and through mechanisms of knowledge spillover and proximity economics.

The chapter intends to make a contribution to the evolutionary economic geography approach focusing on the role of institutions and modes of institutional change in path-dependent processes of innovation, knowledge accumulation and competence building in economic systems. It is organized as follows: section 2 takes a closer look at the connection between institutions, institutional change and innovation on the basis of evolutionary approaches focusing on novelty and using institutions from a macro-level perspective to explain rate and direction of innovation in long-term economic change. In section 3 path plasticity and its relation to institutional change and the role of geography are discussed. In section 4 empirical evidence is provided by exploring the evolution of the German software industry. Despite comparative disadvantages caused by the established institutional setting of the national innovation system, a sub-sector of this industry – *customized business software* – was able to become internationally competitive. The customized business software industry is an example of innovation and successful change in what is described as non-favourable institutional settings. Section 5 draws conclusions and discusses challenges for further research in evolutionary economic geography.

2. Path dependence and co-evolution of institutions and innovation

Institutional analysis is only loosely related to theories of economic evolution (Essletzbichler and Rigby 2007, and Chapter 2 in this book). Nevertheless, in recent years, the need to bring economic institutions into evolutionary theory, and more generally, the need to connect institutional economics with evolutionary economics for the advancement of the theory is highlighted (Nelson, 2002; Pelikan, 2003).

Path dependency is a central concept in both evolutionary and institutional economics, although evolutionary and institutional economics differ in their perspectives on the connection between institution and innovation (Boschma and Frenken, 2006).[1] The focal orientation of evolutionary economics has been on technological change and the role played by institutional factors in the selection and establishment of technological trajectories and the creation of new technologies (Dosi, 1982; Dosi et al., 1988; Nelson and Winter, 1982). Evolutionary economists see the primary function of institutions in moulding the technologies used by a society or shaping technological change itself. The orientation of new institutional economics favours the study of the set of factors that mould and define human interactions, both within and between organizations.

Explaining the ways in which institutions and institutional change affect the performance of economics over time is the broader and central objective of institutional economics. The concept of path dependence has also been established as a key concept in the discussion about institutional change. Dynamic increasing returns, sunk costs, dynamic learning effects, coordination effects, and self-reinforcing expectations are the main mechanisms that lead to path dependence in both technological and institutional change (Arthur, 1989; David, 1993; North, 1990).

Within economic geography, institutional approaches do not constitute a unified paradigm (Essletzbichler and Rigby, 2007, and Chapter 2 in this book).[2] Both institutional and evolutionary approaches have found a way to explain economic development and the uneven distribution of innovation at multiple territorial levels. In the following, a narrow understanding of institutions is proposed by defining them as formal and informal rules guiding actors' perceptions and activities. Institutions should be differentiated from (non-market) organizations (North, 1990). Institutions as the 'rules of the game' are composed of what Scott (2001) named the 'three pillars': the regulative, normative and cultural/cognitive. These depend on different bases of compliance, evoke different logics of actions and offer multi-level bases of legitimacy (Scott, 2001).

For long-term dynamics of economic systems, the relationships between institutions, institutional change and innovation play an important role. Evolutionary approaches concerned with long-run economic development place emphasis on institutions or institutional arrangements and their co-evolution with innovation as important social phenomena on the meso or macro level to explain path-dependent developments and disparities in the rate and direction of innovation performance. They provide insights into the relationship between institutions and innovation leading to path-dependent processes at different spatial levels.

Institutions and innovation in a multi-level view

Innovation activities are distributed very unevenly in space both within and between national economics. Despite the ongoing globalization of the economy, it has become increasingly apparent that there are distinct differences between nations in rates and types of industrial innovation and the variation of innovative sectors contributing to economic performance. International empirical comparisons that use indicators like R&D, patents, export specialization or international trade flows, underline the relative stability of specific innovation profiles and the comparative innovation strength of national economies over time. Furthermore, national specificities in production and trade correspond with distinct differences in the national knowledge base (Archibugi and Pianta, 1992; Guerrieri, 1999; Montobbio, 2004). It seems that the institutions that support the innovation systems remain country-specific, even the systems themselves are becoming internationalized and more intertwined (Carlsson, 2006).

Approaches dealing with the interrelatedness of institutional settings, innovation and competence building over time, such as the systems of innovation (*national*, *regional*, *sectoral*), and concepts in the political economy, like the 'varieties of capitalism' (VoC), place emphasis on the stabilizing function of institutions in the connection between institutions and innovation. These strands of literature, even though developing relatively independently and focusing on different analysing units, have in common the view of the co-evolutionary nature of innovation sometimes more implicitly than explicitly. It is argued that institutions contribute to path dependence of existing systems by reducing uncertainty in innovation processes. The important role of formal and informal institutions is rooted in their shaping of individual and collective learning processes, seen as the foundation of the innovative outcome on the micro level. They operate as selection mechanisms on different levels by setting incentives and constraints. Doing this, institutions have a substantial impact on both the support and restriction of various types of innovation and future learning, hence contributing to specialized knowledge accumulation and competence building over time (Lundvall et al., 2002). The strands of work as aforementioned mainly highlight the selective and retentive impacts of institutional settings in innovation processes contributing to path dependency as discussed in the following.

Responsible for the distinct differences in the institutional structure at the national level is its co-evolution process with the production structure of innovation systems. The interdependence of economic structure and institutions as well as their mutual reinforcement over time determine the modes and outcomes of production and learning (Edquist, 2005; Lundvall et al., 2002; Lundvall and Maskell, 2000). The evolution of the economic structure determines the evolution of the institutional set-up and vice versa. As an outcome of the dynamic interplay of these two dimensions over time, the systemic contexts differ. In reverse, this also provides the explanation for country-specific performance and specialization.

The question of how these co-evolution processes unfold is hardly worked out in detail and the role of institutional change has not been satisfactorily addressed. The majority of studies analysing the relationship between innovation and institutions assume that institutions might be slow in adapting to changes in economic structure. Because of the complex co-evolution processes of economic structure and institutions rooted in history, the rather implicit than explicit basic assumption is that the institutional set-up of countries is relatively stable (Nelson, 1993; Pavitt, 1998). It is assumed that only feedback from radical and basic innovations has the potential to substantially change the institutional setting of national economies (Freeman and Perez, 1988). But such radical innovations tend to remain outside the dominant development path.

Research on sectoral systems of innovation also emanates from the relative stability of national institutional settings that influence innovation processes and trajectories of sectors by providing tangible and intangible resources. Industrial sectors tend to vary systematically with regard to their knowledge bases, knowledge processes and associated sector-specific institutions. According to the sector considered, different sets of actors and institutions have an effect on innovation and economic performance (Malerba, 2006; Malerba and Orsenigo, 2000). In turn, that may explain why diverse institutional settings are co-existing within a national institutional framework. On the other hand, it has become obvious that the international performance of countries in a particular sector is mediated by national and regional institutions and non-firm organizations (Montobbio, 2004). Even though the dimensions of sectoral systems are not necessarily national, but local, national or global as well, it is assumed that national institutions and organizations may in the long run attract those industries most compatible with them (Malerba, 2006).

An explanation for the relative stability of institutional arrangements at the national level compared to the change of sector specific institutions is provided by institutional approaches like the varieties of capitalism (VoC) (Hall and Soskice, 2001), national business systems (Whitley, 2000) or social systems of innovation and production (Amable, 2000).[3] They deliver insights into two mechanisms: institutional coherence and institutional complementarity, both contributing to the slow change and relative *stability of institutional settings at the national level*. Compared with studies of national innovation systems, they are characterized by a more elaborate institutional analysis and differentiation of institutional arrangements. The varieties of capitalism approach treats the respective objectives not as a random collection of institutions, but rather as a pattern of interconnected relationships between the different elements of the institutional structures, which as a result define its coherence (Amable, 2000). Emphasis is placed on the complementary nature of institutional configurations within the economy, which make some institutions more efficient through their interaction with others. It is assumed that the general logic of coordination is probably similar across different key institutional configurations (Casper et al., 2005). Institutional complementarity, which links together different institutions and modes of organization in a certain architecture, contributes to coherence. While the innovation system approaches place emphasis on the differences in the institutional endowment that explain why it is difficult to change a path, the VoC approach highlights particularly the mechanisms of institutional complementarity and coherence.

The question of why in the same national institutional framework several regional innovation systems can emerge and how these are interlinked and intersect with the national one is paid little attention in the literature strand of the VoC approach. Economic geography provides a richness of theoretical and empirical contributions that underline the region as an important organizational level for a variety of processes and territorial-specific selection processes that close down varieties by contributing to path dependency. Agglomeration and localization economies and dynamic localized learning processes are assessed as the main mechanisms leading to region-specific characteristics and localized capability building. Untraded interdependencies and idiosyncratic context conditions evolve over time, fostered by knowledge spillovers and proximity economies that are especially important for tacit knowledge transfer (Cooke, 2005; Gertler, 2003; Malmberg and Maskell, Chapter 18 in this book; Storper, 1997). These mechanisms and processes give explanations for institutional variation at the regional level and the distinct regional institutional endowment. The question of how institutional variation and institutional change at the national and regional level is interrelated and, in turn, how these processes affect innovation at the firm and the sectoral level in a path is in many respects an open one.

Interplay between institutions, competencies and demand

The strands of institutional approaches with different analytical levels emphasize the co-evolution of institutions and innovation over time and provide complementary insights into path dependent processes of economic systems. Especially the interplay between institutional settings, competencies and demand seems to be constitutive for the path-dependent development of innovation systems. The national innovation system concept emphasizes the production structure and user–producer relationship for innovative developments and competence building over time. Research on sectoral systems of innovation shows that innovation processes are determined by sector-specific institutions and in addition it has become obvious that national and regional institutional configurations mediate innovation processes of industrial sectors. Studies based on VoC tend to neglect these influences and pay little attention to the influence of home markets and the quality of the demand side for the emergence and diffusion of innovation. Focusing on institutional configurations at the macro level to explain the behaviour of agents in innovation processes at the micro level, this strand of literature goes beyond the national system of innovation approach in two areas: the development of institutional typologies based on a more systematic and comparative institutional analysis and the consideration of the active role of agency.[4] Research based on VoC succeeds in specifying institutional configurations and their impacts in terms of the room for strategic choice left to firms. The complex institutional configurations, their complementarity as well as their interconnectivity provide comparative institutional advantages for agents. Comparative studies provide empirical evidence that firms are able to use strategically named 'comparative institutional advantages' in developing innovation. These selection processes gradually result in specific knowledge accumulation and competence building of firms, indicated by different national innovation profiles. However, the approach also has some major limitations because the mainly used institutional typology of coordinated and liberal market economies appears robust for some countries – particularly for the US and Germany – while for other countries this appears weaker. Furthermore the co-existence of region-specific institutional settings within a wider national institutional setting is not taken into account.

In sum, the analysed approaches complement each other in their focus on institutions and the way these reproduce stable patterns of behaviour. The complex interplay of different institutional configurations that has evolved and the pattern of interaction between complementary institutions are resulting in a dominant path of national innovation systems in a particular space–time context. Additionally, they share the view of the relative stability of institutional arrangements at the national level compared to the institutional variations and change at the sectoral and regional level. In path-dependent developments, primarily the stabilization and selection function is highlighted by setting incentives for knowledge accumulation and specific competence building – the basis for innovation. There is a lack of discussion in the reviewed strands with regard to institutional change and institutional dynamics within path-dependent developments. Additionally, little is said about how institutional dynamics and institutional change are related to positive and negative lock-in effects of path-dependent developments. Processes of institutional change are mainly seen as either incremental, leading to continuity of the present technological path, or as abrupt and disruptive, leading to the breakdown and replacement of institutional settings. Innovation creates the need to break from the established institutions when new knowledge bases are incompatible with the dominant institutional configurations of a path. This dichotomy underestimates institutional dynamics and institutional change within path-dependent developments that can contribute to positive unforeseen feedback or unexpected impacts over time. The plasticity of institutions and institutional arrangements contributes particularly to 'path plasticity' that will be outlined in the following section.

3. Path plasticity and institutional change

Processes and modes of institutional change within a given path are still insufficiently understood. Apparently the institutional dynamics within a well-established technological and institutional development path is an important condition for avoiding undesired effects of stability, or reducing what Martin (2006) describes as 'negative lock-in effects'. Negative lock-in effects emerge when processes, structures and configurations built up as a result of positive 'lock-in' become a source of increasing rigidity and inflexibility (Martin, 2006; Martin and Sunley, 2006, p. 415, and Chapters 3 and 4 in this book). Garud and Karnøe (2001) place emphasis on the strategic, deliberate and mindful *action* of actors that enable the break-out of path dependence and institutional stability. They focus on the micro level in which powerful actors can strategically create new institutions, push innovation and create new paths.

The notion of 'path plasticity' does not question path dependency or the option of deliberate path creation. It is introduced here to make the point that there is 'plasticity' within a well-established institutional setting of technological development paths, which enables innovation even with only a minor degree of compatibility to come into being without necessarily breaking out of a path. Path plasticity is used here to set the focus on the continuity of dynamic change, while path creation puts emphasis on the 'creative destruction' and the mode of disruptive institutional change.

The term 'plasticity' was first introduced by Alchian and Woodward (1988, p. 69), who used plasticity in the theory of economic organization to show that resources and investments are plastic. They indicated that there is a wide range of discretionary, legitimate decisions the user may choose, thus claiming that this is underestimated in transaction cost theory, especially with regard to the moral hazard and principal–agent problem. Zysman (1994, p. 261) used the term 'social plasticity' to make the case that technology is a socially created constraint. Thus strategies and tactics for approaching technological problems will vary from place to place. Strambach and Storz (2008) argue there are two sources of plasticity of innovation systems for which they use the notions numerical and functional plasticity. Systems are configurated by a multitude of elements. The notion of numerical plasticity is understood as the sum of these elements in relation to the whole system, and is seen as a precondition for functional plasticity. The latter refers to the way configurations can be moulded to produce new uses and can be adapted to new functions or purposes. According to Strambach and Storz (2008) there is numerical and functional plasticity in any given innovation system. They consider the institutional and structural variety proved by the literature on regional and sectoral innovation systems in recent years as evidence for the high degree of numerical plasticity in national innovation systems.

In the following, the chapter builds on this differentiation and expands more specifically on the plasticity of institutions and institutional arrangements and their contribution to the plasticity of paths. Geography does play an important role in the exploration and exploitation of institutional plasticity through proximity effects and place-specific characteristics.

Plasticity of institutions

Besides their action-guiding function highlighted by the work on systems of innovation, institutions simultaneously act as enablers, while actors can use institutions as toolkits in a myriad of ways to solve innovative problems. They are able to recombine and convert or reinterpret institutions for their new objectives or transfer institutions to different contexts. By doing this, actors shape and form institutions and are themselves becoming influenced by the institutions. Giddens (1984) points to this fact with the notion of 'duality of structures'. The plasticity of institutions results from the interpretative flexibility of their meaning. The selection impact of institutions based on the establishment of incentives and constraints is dependent on the assessment of actors. The space for interpretative flexibility of institutions differs with regard to different kinds of institution and is defined by their sanctions for deviation.

Formal, regulative institutions such as laws or standards provide little room for the flexible interpretation of their meaning compared to informal, normative rules like values or norms. The deviation from the latter is associated with social sanctions, whereas the deviation from regulative institutions is mostly legally sanctioned, which reduces the room for interpretative flexibility, yet by no means removes it totally. Sanctions generally enhance the probability that actors commit and follow one dominant interpretation, but others do not cease to exist.

One important feature of place specificities is the intersection of multiple institutional configurations, which produces a rich environment for variation. The overlapping of various firm- and industry-specific institutions and their intersection with national institutional settings provides a repertoire of already existing institutional compositions in idiosyncratic context conditions evolved over time. Actors are able to use these pre-existing institutions for the creation of new solutions in ways that may lead to evolutionary change. Geography fosters processes of exploring and exploiting institutional plasticity. Spatial proximity often combined with other types of proximity (Boschma, 2005) and localized learning of individual and collective agents (Malmberg and Maskell, 2006) facilitate both the exploration of the interpretative scope of institutions as well as the closing down of the varieties of meanings and the coming through of an interpretation as the dominant one. Whereas particularly social and cognitive proximity (Boschma, 2005) may facilitate the consolidation of a dominant meaning among actors through shared understanding, unanticipated encounters and neighbourhood effects (Malmberg and Maskell, 2006) may lead to opening the space for the interpretative flexibility of an institution by getting to know a variety of different actors' meanings. Institutional forms arising from such processes need not necessarily be completely new, they are rather novel combinations of earlier institutional components. These are created together by communities of actors, reconfigured and combined into various hybrid forms to serve new or modified goals. In localized learning processes, aspects of meaning and legitimacy from earlier institutions are transferred to the novel combinations that in turn facilitate their coming through. The exploration and exploitation processes of institutional plasticity to achieve new purposes start at the micro level, but contribute to institutional change and institutional dynamics within path-dependent processes. Plasticity of path does not only exist at the micro level through the interpretative flexibility of institutions, it is also based on the elastic stretch of institutional configurations at the macro level.

Plasticity of institutional configuration

Institutional complementarities and coherence are obviously important mechanisms for the stability of path-dependent developments by generating disincentives to radical change. But it is often neglected that the composition of institutional configurations is not static, but rather simultaneously providing a flexible scope for change. Amable (2000) points to the institutional hierarchy, the relative importance of one or a few institutions for the coherence and dynamic of the institutional architecture as such. Taking into consideration the institutional hierarchy, the transformation of one institution within institutional configurations need not necessarily destabilize the coherence of a whole architecture. That may explain why several regional innovation systems with region-specific institutions may exist and can be absorbed by the key institutional arrangements at the national level. But on the other hand, institutional change in one sphere can increase pressure and have a snowball effect on complementary institutions to change gradually. The accumulation of incremental change over time may lead to what Streeck and Thelen (2005) describe as 'transformative change'. Institutional complementarity therefore plays an ambiguous role by contributing to the stability of path-dependent developments and, at the same time, feeding into the plasticity of paths.

Streeck and Thelen (2005) argue that it is not clear at all whether the two basic models of institutional change – incremental and disruptive change – exhaust the possibilities, or even that they capture the most important ways in which institutions evolve over time. They refer to 'transformative change' of institutions as a type of change contributing to capturing current developments in the economy of modern capitalism. Different modes of gradual transformation of institutions such as displacement, layering, drifting or conversion and exhaustion are explored, resulting in institutional discontinuity.

The perspective of path plasticity lays emphasis on the interpretative flexibility at the micro level, which enables the slow evolution of institutions and the elasticity of institutional arrangements at the macro level, supporting institutional change and institutional dynamics. Plasticity allows institutional variations, the attachment of new elements to existing institutions, the slow rise of peripheral meanings to dominant institutions and their conversion by the redeployment of old institutions to new purposes (Streeck and Thelen, 2005). Processes and modes of institutional change within a given institutional setting of a technological development path make it possible that initially incompatible innovations may over time evolve into innovation systems. In the following section empirical evidence is provided by the evolution of the German business software industry, which has been developed despite the institutional disadvantages ascribed and caused by the dominant institutional configurations of the national innovation system.

4. The evolution of the software industry in Germany

The evolution of software as an independent industry is a relatively recent phenomenon. As a cross-sector technology, software has become a major integral part in processes, products and services. Software development and software services are provided for the market by firms in the primary software branch, but also by corporations in user branches, the so-called 'secondary industry branches', such as mechanical engineering, vehicle construction or telecommunications (Friedewald et al., 2001). A pronounced secondary software industry is a special characteristic of the German innovation system. Customarily such firms are differentiated with regard to their performance spectrum into providers of standard software, called *packaged software*, and providers of *customized software*.

The software industry belongs to knowledge-intensive business services and as such it has sector-specific institutional characteristics:

- The intensive user–producer interaction and communication linked with the production of software goods. This is particularly necessary during the creation of customized software and IT services. Users are directly involved in the value added activities, because processes of knowledge exchange and knowledge sharing are necessary from both sides and determine the quality of the software good.

- Project organization is the dominant form of work organization in general and it is highly significant for the integration of external knowledge for the many small- or medium-sized companies. As in other European countries in quantitative terms the German software industry is dominated by small firms.
- High coordination costs as a result of the integration of myriad knowledge sources in the product and service development – the key function of formal and informal network relations, the role of references and reputation as coordinating mechanism for transactions and interactions.

The markets for knowledge-intensive business services are characterized by the high volatility, uncertainty and ambiguity of projects, and marked heterogeneity of the competencies involved. The low formal constraints on market access allow for fast market entries, which, however, are accompanied by a high ratio of market exits.

Path dependency – the software industry in the German national innovation system
In international terms the software market is dominated by US-based firms. The German innovation system did not succeed in establishing itself in the new knowledge-intensive service industries, like the software industry. According to Meyer-Krahmer (2001, p. 208) the information technology (IT) industry is characterized by specialization disadvantages and is lagging far behind the dynamism found in the US. Even though software production in Germany increased between 2003 and 2006 with a rate of 13.6 per cent, it did not reach the dynamic level of Europe (16.4 per cent) in these years (EITO, 2006, pp. 199–207). The international performance of countries in a particular sector is mediated by national and regional institutional settings. The complementary institutional arrangements of the dominant technology-oriented development path of the German innovation system are assessed as unfavourable for the emergence and success of the software sector (BMBF, 1999, 2006, 2007).

The main feature of the long-term development path of Germany's national innovation system is its distinct industry-based innovation profile. In contrast to most other larger economies, the R&D-intensive manufacturing industries maintained their macroeconomic weight within the national economy in the last ten years. These industries have an aggregate share of 39 per cent of the added value of the overall economy. Nearly 60 per cent of the country's total amount of export is R&D-intensive technology goods (Table 19.1). In 2004 Germany was the largest exporter of technology goods with a share of 14 per cent of world trade, compared to the USA with 13.2 per cent and Japan with 10.7 per cent (BMBF, 2007).

In international comparisons there is a common distinction at the level of industrial sectors in a horizontal dimension that classified industries as high-tech, medium high-tech or advanced, and as low-tech industries. Of course innovation happens in all kinds of industry, but it is suggested that innovation processes and sector-specific knowledge bases have distinct differences. Radical innovations involving large expenditures on R&D and analytical knowledge base are more common for high-tech industries, while the development of incremental innovations based on cumulative knowledge bases

Table 19.1 Trade in R&D-intensive goods, Germany, 2005

	Export	Import	Trade balance	Share of total exports	Share of total imports	Share of trade balance
	in billion €			in %		
R&D-intensive goods	428.3	264	164.3	59.9	55.6	68.4
High-tech	96.7	95.8	0.9	13.5	20.2	0.4
Aircraft and spacecraft	21.4	21.2	0.2	3	4.5	0.1
Telecommunications and sound-apparatus and equipment	20	18.3	1.7	2.8	3.9	0.7
Automatic data-processing machines	19.8	26.9	−7.1	2.8	5.7	−2.9
Electronics	12.6	13.7	−1.1	1.8	2.9	−0.5
Optics, medical technology,	12.3	5.5	6.8	1.7	1.2	2.8
Medicinal and pharmaceutical products	7	7.7	−0.4	1	1.6	−0.3
Radioactive goods, nuclear reactors	2.1	1.6	0.4	0.3	0.3	0.2
Pest control, pest management, crops science	1.5	0.8	0.7	0.2	0.2	0.3
Medium high-tech	328.6	161.4	161.4	46	35.2	67.1
Motor vehicles, parts and accessories	147.1	85.2	85.2	20.6	13.1	35.4
Machinery	60.6	38.2	38.2	8.5	4.7	15.9
Chemical materials and products	49.1	16.9	16.9	6.9	6.8	7
Office machines and electrical machinery	24	7.7	7.7	3.4	3.4	3.2
Medicinal	23.6	6.6	6.6	3.3	3.6	2.7
Medical technology, instruments, optical equipment	13.3	6.6	6.6	1.9	1.4	2.8
Rubber manufactures	5.4	0.6	0.6	0.8	1	0.3
Radio and television technique	2.9	−2.3	−2.3	0.4	1.1	−0.9
Railway vehicles	2.7	1.8	1.8	0.4	0.2	0.8
Non-R&D-intensive goods	286.6	210.6	76	40.1	44.4	31.6
TOTAL manufacturing	715	474.6	240.4	100	100	100

Source: BMBF (2007), p. 34, adapted.

characterizes medium high-tech industries. In the latter the German system has its strength.

Noticeable comparative advantages exist in medium high-tech, and also specific advanced technology fields of the core industries – like vehicle construction, mechanical engineering, electrical engineering and the chemical industry. Patents, particularly the RPA (relative patent activities) indicator, are used to compare the new knowledge production of different economies.

Inter-country comparison shows that Germany not only has the highest specialization figure in advanced technology fields, but also that this specialization expanded between 1991 and 2004 (Table 19.2).[5] The number of patents, export specialization and an above-average share of world trade in the automobile manufacturing, mechanical engineering, and electrical engineering sectors demonstrate Germany's strength in application-oriented advanced technology innovation.

Particularly research based on the VoC approach has elaborated the role played by the key institutional configurations in the development of the specific innovation profile. The interdependence and complementary nature of the dominant institutional settings – the coordinated system of industrial relationships, the labour market institutions, the closely knit education and training system and the specific financial institutions – reinforce the international competitiveness of the core industries by contributing to the accumulation of specific competencies (Naschold, 1997; Soskice, 1999; Streeck, 1997). The path-dependent specialization pattern of the national innovation system based on the co-evolution of the technological and institutional development path has not changed much in the last decade.

The downside of this specialization pattern is the associated structural weakness of the system – namely the comparative disadvantages in top grade technologies characterized by more radical innovation processes and knowledge-intensive service industries. High-technology goods are of only minor importance in exports (Table 19.1) and the patent applications in high-technologies are still underrepresented in international comparison. The share of R&D expenditures in services and high-technology sectors is far below the OECD average both in 1991 and in 2003 (Table 19.2). In services the share of R&D expenditures in 2003 was by average three times higher in the OECD than in Germany. Only in one high-tech industrial sector – medical precision and optical instruments – did German R&D expenditures reach the OECD average (Table 19.2). As underlined by empirical indicators, the German innovation system did not succeed in establishing itself in the new science-based industries characterized by radical innovation and knowledge-intensive service industries, like the software industry.

The interplay between institutional settings, competencies and the quality of demand was identified as constitutive for the path-dependent development of innovation systems. How this co-evolving interplay affects the development of the German software industry will now be analysed in more detail.

Industrial organization The comparatively high degree of vertical integration of business-related service functions in international terms is an essential element of the competitiveness of the core industrial sectors of the national innovation system. This kind of institutional organization fosters the *competence building* that produces complex, systemic technology- and service-intensive goods. The industrial firms of the key sectors

Table 19.2 Structure of R&D expenditures by sectors, Germany and OECD, 1991, 1997, 2003

	OECD			Germany			Germany's share in OECD		
	1991	1997	2003	1991	1997	2003	1991	1997	2003
		in %			in %			in %	
Services	14.3	14	24.1	2.4	5.4	8.5	1.8	3.3	3.1
High-tech industry	42.3	44.7	39.2	34.9	33.9	30.2	8.8	6.5	6.8
Pharmaceuticals	6.8	8.1	9	5.6	6.5	8	8.8	6.9	7.9
Office, accounting and computing machinery	7.9	8.2	4.8	4.9	2.3	1.4	6.6	2.4	2.5
Radio, television and communication equipment	11.6	14.4	12.8	14.5	11.3	8.7	13.4	6.7	6
Medical, precision and optical instruments	5.6	6.4	7	1.7	5.2	7.1	3.2	6.9	8.9
Aircraft and spacecraft	10.3	7.5	5.7	8.1	8.5	5.1	8.5	9.8	7.9
Medium high-tech industry	30.5	29.1	26.2	53	51.8	53.6	18.6	15.2	18.1
Chemicals excluding pharmaceuticals	8.4	6.5	5.1	14.2	12.2	8.7	18.2	16.2	15.1
Machinery and equipment, n.e.c.	5.5	5.9	5.4	10.5	11	9.9	20.6	16	16.3
Electrical machinery and apparatus, n.e.c.	5.5	4.4	3	10.3	3	2.8	19.9	5.9	8.3
Motor vehicles, trailers and semi-trailers	10.8	11.9	12.2	17.5	24.2	31.8	17.4	17.4	23.2
Railroad equipment and transport equipment n.e.c.	0.3	0.4	0.5	0.4	1.3	0.4	12.4	25.8	6.9
Medium-tech	11.6	6.1	4.6	6.3	5.5	5	8	7.7	9.5
Low-tech technology		3.8	4	2	2.2	2.2		4.9	4.8
Other manufacturing	1.9	2.3	1.8	1.6	1.1	0.5	9.1	4.3	2.7
TOTAL	100	100	100	100	100	100	10.7	8.6	8.9

Source: BMBF (2007), p. 64, adapted.

use product-attendant services for product differentiation and strengthening their market position. It is shown that the intensification and profiling of product-oriented services are regarded as core competencies for establishing comparative advantages (Hornschild et al., 2003; Stille, 2003).

At the same time, the high degree of vertical integration is one factor that has hindered the development of new, more challenging, markets in the knowledge-intensive services sectors (BMBF, 2006; Strambach, 1997, 2002a). The development of software in Germany mainly takes place in the secondary industry branches, such as mechanical engineering, vehicle construction or telecommunications, and is oriented towards the customers of these industries (Friedewald et al., 2001). It is assumed that 80 per cent of German software engineers are employed in the secondary software sector (Broy et al. 2006). The long-established, above-average ability of industrial firms in Germany to practise mainly intra-firm software development has delayed the formation of a clearly competitive and specialized software branch like those in other countries (Casper et al., 1999; Lehrer, 2000, 2006).

Demand The high demand of the industrial firms of the core industrial sectors and the large SME sector — often called the 'Mittelstand' — for individually adapted software solutions for their production and business processes, has contributed to the dominance of customized business software development. The production of *customized business software* is disadvantageous, because its use is limited to a particular enterprise, and the ability to re-use parts of software developed for earlier customers is very small (Holl et al., 2006). The production of software primarily for individual customers and the resulting lack of uniform standards impede the development of economies of scale in the software industry. Further, these factors have helped to orient the business services sector towards the domestic market that increasingly faces pressure from foreign competitors (BITKOM, 2007). In addition, the negotiation-oriented organization of industrial firms does not allow the services firms to position themselves rapidly in the market and to reorganize in a flexible way.

Labour market The regulated labour market and the labour market institutions that characterize the German innovation system support knowledge accumulation and the competence building within firms. Similarly, it is evident that the lack of flexible labour market institutions, like those typical of LMEs such as the USA and the UK, is a major competitive disadvantage for the software industry with its industry-specific institutions. The labour market institutions focusing on long-term employment prevent firms from adapting quickly and make the transfer of experience-based and applied knowledge difficult. However, this mainly tacit knowledge bound up in employees and networks is particularly valuable for the project-based software industry and its innovation processes. In international comparisons especially, the rigid regulated working hours and the regulations relating to hiring and firing of employees are considered to be a major factor accounting for the lack of flexibility of the software firms (BITKOM, 2007; Djankov et al., 2006).

Additionally, the legal requirement that employees must be represented in the decision-making negotiations, and their co-determination there, make these processes very time consuming. Organizational change and the introduction of organizational innovations that are decisive for fast adaptation to the dynamic changing markets of the software industry are hampered (Strambach, 2002b). Casper and Vitols (2006) provide empirical evidence that the institutional arrangements of the German coordinated market economy are inappropriate for supporting the competence building of project-based software firms. The institutional regulations cause disadvantages for German firms in the rapidly changing markets of high-technology and hamper the necessary fast market entry and exit.

Education system The closely knit connections between industry and the university system and the interconnected further education and occupational training systems are considered to be a major strength of the German innovation system (Naschold, 1997; Soskice, 1999; Streeck, 1997). Overlapping, interrelated qualification structures of skilled workers, technicians and engineers have been built up by the dual training system, which ensures the transfer of technology-oriented knowledge and technical capabilities to the production system. For the software industry, such interlocking structures are missing (Broy et al., 2006).

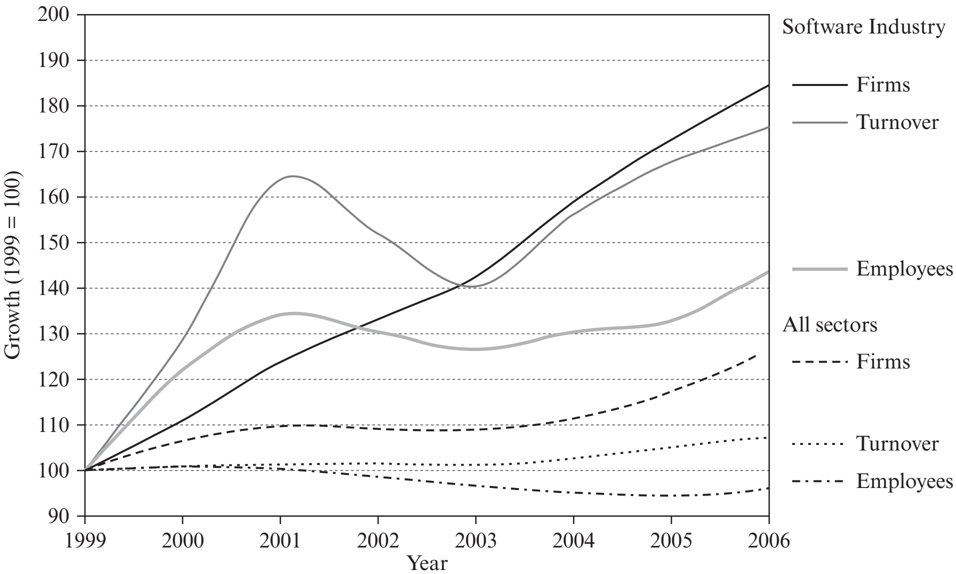
Furthermore, it is emphasized that the failure of the standardized education system to adjust to the dynamic changes and requirements of the software branch was a factor that slowed the growth of this industry. The complex voting processes within the corporate structures meant that it took a long time for changes in qualification requirements to be carried over into new job descriptions by way of new education regulations. The orientation of the education system towards clearly defined areas of activity and job descriptions means that there are no interdisciplinary application-oriented training courses and, subsequently, that there is a shortage of qualified employees (BITKOM, 2007; BMBF, 2007; Friedewald et al., 2001).

In sum, it can be stated for Germany that the national institutional settings and the co-evolved specific competencies in certain technology fields have resulted in specialization of the production structure in the core industrial sectors that constitutes a disadvantage for the genesis of strong, growing, new high-tech industries like the software industry.

Path plasticity: the customized business software in Germany

Despite the unfavourable institutional environment caused by the institutional arrangements of the dominant path of the German innovation system for the software industry, a sub-sector – *customized business software* – has been able to position itself internationally on the world market.[6] The growth dynamic of the application software in Germany with a rate of 13.6 per cent outperformed that of the European Union in the years 2003 to 2006 although there is a lack of system software (EITO, 2006, pp. 199–207) (Figure 19.1).

In 2006 the broader sector of computer and related services (NACE 72) consisted of 65,440 firms, over half of the firms focusing on software development. In absolute terms, with 402,311 employees the computer and related services industry has nearly as many employees as the chemical industry – one of Germany's core industrial branches. Around a third work in the software industry, which is highly concentrated in Germany with only a few large international companies and a large number of small and very small, mostly single, firms. The structural survey of the National Statistical Office for 2005 shows that 87 per cent of software companies employ up to nine employees. Only around 6 per cent of the firms have more than 20 employees, but these are producing 80 per cent of the turnover of the whole branch. Since the mid-1990s the software branch has been characterized by a dynamic continuous growth in software companies. The number increased from 5700 firms in 1994 to 19,000 in 1999 and to 35,700 in 2006 (see Figure 19.1 and



Sources: Own calculations based on employment statistic *Bundesagentur für Arbeit* and turnover tax statistic *Statistisches Bundesamt* (several volumes).[7]

Figure 19.1 Development of the German software industry 1999–2006

Appendix Table 19A.1). These firms generated a turnover of 24.1 billion euros in 2006. Compared to the overall economy the software industry proved to be highly dynamic in the years 1999 to 2006 with an impressive increase in firms (+84.5) and turnover (+75.5) (see Table 19A.1). Between 1999 and 2007 a total of 112,712 (+64.15 per cent) new jobs were generated, with 23 per cent of these situated in Baden-Württemberg, which is the leading location of the German software industry. In this region 16 per cent of the German software firms, 23 per cent of the turnover and 24 per cent of the employees are concentrated. Baden-Württemberg has the highest number of employees and firms in the software industry in absolute terms. Global players as SAP AG (Walldorf), SAS (Heidelberg), IBM Global Services Germany (Stuttgart) or Hewlett-Packard (Böblingen) have their location there. SAP is a world market leader in business application software (BITKOM, 2007).[8] In recent years the OECD (2006, p. 68) stated that Germany is already a 'leading exporter of software goods'. The international significance of its customized business software is surprising, because – in contrast to the USA – there are no large hardware producers in Germany able to function as carriers for the software.

The concept of path plasticity may be able to provide explanations connected with the co-evolution of institutions, competencies and demand. The chapter argues that the *plasticity of institutions and institutional configurations* has left actors room for strategic choice and therefore makes an important contribution to 'path plasticity'. The institutional plasticity of the dominant development path of the German national innovation system and the ways German software firms have strategically used the room for manoeuvre to achieve their innovative developments is discussed in the following.

Institutions In the service industries the labour markets in Germany are more open than those of the key industrial sectors. They have a lower degree of trade union organization and, up to the foundation of Verdi in 2001, service industries were fragmented with many single trade unions. Flexible working organizations using freelancers, part-time workers and personnel leasing are common in service industries. In particular the large manufacturing firms of the key industrial sectors have benefited by spinning off their IT services, for instance Debis System House (formerly integrated in Daimler) and Siemens Business Systems. By using this strategy the large corporations have enhanced their flexibility and avoided the rigid working hour regulations and the collective negotiations legally required in manufacturing industries. As empirical studies have shown, the mainly small project-based firms of the primary software branch use a large amount of freelancing (Bertschek et al., 2006). This kind of work organization allows them to adapt more quickly to dynamic market changes. At the same time, flexible personnel management ensures that the firm accumulates internal knowledge through the long-term employment of a core workforce.

Additionally, for small firms of up to five employees the main regulations do not apply, so that temporary employment contracts, or flexibility options with regard to protection from being fired, are legally possible. Such rather peripheral institutions within the dominant labour market arrangements are minor considerations in the discussions of the German national innovation system, but they are very important for the competitiveness of the mainly small software firms. Institutional change is occurring, triggered by the federal association of the ICT industry 'BITKOM' which is currently demanding these flexibility options to be extended to firms of up to 20 employees (BITKOM, 2007).

German software providers also fall back on the institutions that are part of the dominant path of the innovation system and are adapting these to the requirements of the new sector. Institutional organization in associations and the cooperation in networks are the two main instruments that are used by the software industry. The common interests of the associations have led to the integration of the heterogeneous but overlapping sub-sectors of the information industry – the telecommunications and media industries – into one single federal industry association BITKOM. This has allowed the actors to reach the critical size to initiate institutional changes within the key institutional configurations of the education system and the labour market to improve their framework conditions. The fragmented associations of the sub-sectors have not ceased to exist but have been complemented through the establishment of new ones at the regional and local levels. Such variety facilitates the exchange of specialized experience-based knowledge between partners operating in similar knowledge domains and therefore having more cognitive proximity. This in turn makes the absorption of knowledge easier (see Boschma, 2005; Nooteboom, 1999). It is beyond the scope of the chapter to analyse this institutional change processes in detail, but it is worth mentioning that the initiatives stem regionally from Baden-Württemberg and the chairman of the first regional business network named Baden-Württemberg Connected (bwcon) became also later the first chairman of the national association BITKOM.

The furthering of collaboration and the cooperation in networks are regarded as institutional advantages of the dominant path of the German innovation system (Casper and Vitols, 2006). Software firms producing customized software are confronted with a double pressure: the need to simultaneously provide highly specialized, knowledge-intensive expertise and comprehensive problem solutions (Strambach and Storz, 2008). By collaboration they are able to deal with the problem of gaining access to a number of heterogeneous knowledge bases and to be able to combine and reconfigure these into composite knowledge products adapted to the specific customer context.

Long-term customer and cooperative relationships also support business software firms in creating the network externalities that are very important for the success of *business software*. The accumulated experience-based application knowledge of the complex customer contexts generates high transaction costs that are wasted by changing the provider. The large international industrial companies in the core sectors are an important interface for the penetration of the customized business software. These relatively new, mostly small, software development firms are unable to do this alone. Innovative, individual problem solutions are applied to the integrated firm-internal production systems by way of long-term customer and cooperative relationships with strong industrial export-oriented customer firms located in many parts of the world. In addition, the internationalization and outsourcing processes of the industrial producers also support the development of innovative software solutions within the context of other firms through their wide networks of suppliers. Although the coupling of software and hardware that plays an important role in the USA is not available to the German software firms, the function of large industrial enterprises as carriers for the product distribution is nevertheless similar. Business software firms have been able to use the supposed competitive disadvantages of the institutional arrangements by converting them into advantages for innovative processes and thus to hold their own in international markets.

Demand The quality of demand is an important precondition for the development of customized business software in Germany. Second to the US, Germany is the country that spends the most on ICT, ahead of the UK and Japan (OECD, 2006). The technically challenging demand for inter-operable software solutions able to communicate with existing systems and applications by both firms of the economically successful manufacturing branches and the large SME sector foster the competencies for producing complex products across systems (see BITKOM, 2007). In addition, the accumulated knowledge gained in the course of adapting existing legacy systems of clients contribute to this competence building. Legacy systems are a special characteristic of enterprises in the secondary software industry in Germany (GfK et al., 2000). Such systems contain applied knowledge that has matured and accumulated over many years. They were developed using what are now outdated methods and, despite the existence of more efficient systems, they have not been replaced because of high costs of change. The adaptation of such complex systems supports the building up of experience-based knowledge by using multifaceted interfaces and technological platforms of different ages that have to be combined and integrated with new innovative systems and technologies. The mostly international operations of the user firms in the key industrial sectors imply a further challenge for the primary software firms developing high-quality innovative and competitive business software.

Competencies The presumed competitive disadvantages, namely the intensive interaction processes with demanding and quality-critical manufacturing customers and the large number of individual software solutions that are being provided play a decisive role for the accumulation of the distinct competencies. The German firms are able to use these competencies favourably in innovation processes. Obviously they are able to find ways of overcoming the two main disadvantages resulting from the production of mainly customized business software. The limited re-use of parts of the developed business software and the high degree of necessary interaction-intensive service activities used to implement the enterprise solution are a major limitation for tradability and achieving economies of scale.

An enabling factor is the transfer and adaptation of experience-based knowledge of engineering and industrial production and the business processes used in the complex production systems of the core manufacturing industries to both the software development and the service process. Software engineering and the so-called 'service engineering' (Streich and Wahl, 2006) are means of enhancing the efficiency of developing customized software products by using modularization and standardization. In turn, modularity and standardization permit a higher degree of labour division both in the software development and in the service process by creating possibilities for externalizing parts to other specialized providers. Particularly in the last years growth has been strong in the IT outsourcing market in Germany and therefore questions have been raised about its laggard position (Lehrer, 2006). The specialization in software architecture and the design of highly complex and comprehensive solutions allow the German software firms to use their competencies in innovation processes favourably. Heinzl and Oberweis (2007) identify the systemic way of thinking and the engineering and process knowledge that have evolved over decades as a sustainable competitive advantage for the German customized business software.

5. Conclusions and challenges for further research

The chapter contributes to the evolutionary economic geography approach by addressing the role of institutions and modes of institutional change in path-dependent processes of innovation, knowledge accumulation and competence building in innovation systems. It argues that there is a lack of discussion regarding institutional dynamics within path-dependent developments. Processes of institutional change are mainly seen either as incremental, leading to continuity of the present technological path, or as abrupt and disruptive, leading to the breakdown and replacement of institutional settings. Especially when new knowledge bases are incompatible with the dominant institutional configurations of a path, it is argued that innovation creates the need to break away from the established institutions.

The evolution of German customized business software points out that radical institutional change is not always necessary for the successful introduction of innovation even in non-favourable institutional settings of the dominant path. The success of this industry is an indication that paths are not coherent in themselves. Actors have harnessed the plasticity of institutional configurations by selecting peripheral and dominant elements to combine and adapt these for their requirements. Plasticity permits institutional variations and the conversion and redeployment of established institutions for new purposes by agents. The customized business software industry chooses hybrid solutions for dealing with disadvantages of the institutional setting. The advantages of the existing institutional arrangements are not replaced totally by the introduction of new ones. By recourse to peripheral elements of the dominant path the firms achieve a higher degree of flexibility and retain established elements of the dominant path to advance knowledge bases. Hence innovation may not be the unavoidable result of 'creative destruction' that underlies the breaking up of institutional stability. Instead it can be the result of 'creative accumulation' (Malerba, 2006, p. 4) and institutional dynamics in path-dependent development. In other words, it can be argued that it is precisely the path-dependent development of the German innovation system that has contributed to the genesis and the success of the customized business software sector.

Evolutionary economic geography could significantly contribute to provide new insights in long-term dynamics of economies in space and time by making institutions an integral part of the analysis. Beyond the use of institutions to explain inertia and stability, there is a need to specify institutions and institutional configurations in a more systematic way and to analyse their impacts on the dynamic interplay between actors and structures of the economic system in time and space. Geography itself does play an important role in the exploration and exploitation of path plasticity through proximity effects on processes of knowledge transfer and creation at the micro level. Additionally, place-specific institutional compositions provide a rich repertoire for variation that can be used by actors to recombine and adapt pre-existing institutional components for new requirements in order to achieve innovative solutions. These change processes may have impacts on macro institutional configurations of a dominant path of innovation systems by enabling the slow evolution of institutions and transformative change. How processes of institutional change and institutional dynamics are interrelated with innovation processes at multi-level spatial scales is an important question with several methodological and conceptual challenges, but also opportunities for the developing evolutionary economic geography.

Notes

1. It is worth to mentioning that the understanding of institutions is also differentiated in the areas of evolutionary and institutional economics.
2. For a broader overview in economic geography see Martin (2000).
3. The varieties of capitalism (VoC), national business systems (NBS) and social systems of production and innovation (SSPI) approaches have strong commonalities even though they differ slightly in their institutional analysis. The following analysis concentrates on the VoC approach.
4. An archetype of distinct institutional frameworks is that of the coordinated market economies (CME), which characterize for example Germany, Sweden or Japan, or the liberal market economies (LME) typical for the US, UK or Canada. These economic systems differ substantially in their organization and the governing of key institutional configurations: the financial system, the industrial relations system, labour markets, the education and training system, and inter-company relations (Hall and Soskice, 2001; Soskice, 1999; Whitley, 2002).
5. In 2004, the RPA value for Germany (+16) was four times higher than the EU-15 average (+4) (BMBF, 2007, p. 45).
6. The chapter is based on research supported by the VW-Foundation under the programme: 'Innovation processes in economy and society', which is gratefully acknowledged. Background for the qualitative analysis is built on empirical interviews with 21 software firms and firms from user branches and intermediary organizations of the IT industry. We are especially grateful for the assistance in the empirical analysis from Benjamin Klement and Konstantin Schneider.
7. Annotations: The software industry corresponds to NACE 72.2 (Rev. 1.1), 'all sectors' corresponds to NACE A-O (Rev. 1.1) and the illustrated employees are employees subject to social security contributions.
8. The terminology is not selective with regard to *customized business software* and the application software. Sometimes they are used synonymously, sometimes customized business software is aggregated under application software. In the following both terms are used synonymously.

References

Alchian, A.A. and S. Woodward (1988), 'The firm is dead, long live the firm: a review of Olliver E. Williamson's *The Economic Institutions of Capitalism*', *Journal of Economic Literature*, **26**, 65–79.

Amable, B. (2000), 'Institutional complementarity and diversity of social systems of innovation and production', *Review of International Political Economy*, **7** (4), 645–87.

Archibugi, D. and M. Pianta (1992), *The Technological Specialization of Advanced Countries: A Report to the EEC on International Science and Technology Activities*, Dordrecht & Boston: Kluwer Academic Publishers; Commission of the European Communities.

Arthur, W.B. (1989), 'Competing technologies, increasing returns, and lock-in by historical events', *The Economic Journal*, **99** (394), 116–31.

Bertschek, I., B. Müller, J. Ohnemus and K. Schleife (2006), *IT-outsourcing, Internationalisierung und Flexible Arbeitsorganisation: Stragien im Zeitalter der Globalisierung: Unternehmensbefragung Herbst 2005 in Baden-Württemberg*, Stuttgart: Fazit Forschung.

BITKOM (ed.) (2007), *Zukunft digitale Wirtschaft. Volkswirtschaftliche Bedeutung der ITK-Wirtschaft. Strategische Wachstumsfelder. Empfehlungen an Politik und Unternehmen in Deutschland* [The Future of Digital Economy. The Economic Significance of the IT-Economy. Strategic Areas of Growth. Recommendations to Politics and Companies in Germany], Berlin: BITKOM.

BMBF (1999), *Bericht zur technologischen Leistungsfähigkeit Deutschlands 1998* [Report on Germany's Technological Competitiveness 1998], Bonn, Berlin: BMBF, Referat Öffentlichkeitsarbeit.

BMBF (2006), *Bericht zur technologischen Leistungsfähigkeit Deutschlands 2006* [Report on Germany's Technological Competitiveness 2006], Bonn, Berlin: BMBF, Referat Öffentlichkeitsarbeit.

BMBF (2007), *Bericht zur technologischen Leistungsfähigkeit Deutschlands 2007* [Report on Germany's Technological Competitiveness 2007], Bonn, Berlin: BMBF, Referat Öffentlichkeitsarbeit.

Boschma, R. (2005), 'Proximity and innovation: a critical assessment', *Regional Studies*, **39** (1), 61–74.

Boschma, R.A. and K. Frenken (2006), 'Why is economic geography not an evolutionary science? Towards an evolutionary economic geography', *Journal of Economic Geography*, **6** (3), 1–30.

Boschma, R.A. and R.L. Martin (2007), 'Editorial: Constructing an evolutionary economic geography', *Journal of Economic Geography*, **7** (5), 537–48.

Broy, M., M. Jarke, M. Nagl and D. Rombach (2006), 'Manifest: strategische bedeutung des software engineering in Deutschland', [Manifest: the strategic significance of software engineering in Germany], *Informatik Spektrum*, **29** (3), 210–21.

Bundesagentur für Arbeit (2007a), *Sozialversicherungspflichtig Beschäftigte nach Wirtschaftsgruppen – Deutschland* [Employees Who Are Subject to Social Insurance According to Economic Groups – Germany], several volumes, Nürnberg: Bundesagentur für Arbeit.

Bundesagentur für Arbeit (2007b), *Sozialversicherungspflichtig Beschäftigte nach Wirtschaftsgruppen – Baden-Württemberg* [Employees Who Are Subject to Social Insurance According to Economic Groups – Baden-Württemberg], several volumes, Nürnberg: Bundesagentur für Arbeit.

Carlsson, B. (2006), 'Internationalization of innovation systems: a survey of the literature', *Research Policy*, **35** (1), 56–67.

Casper, S. and S. Vitols (2006), 'Managing competencies within entrepreneurial technologies: a comparative institutional analysis of software firms in Germany and the UK', in M. Miozzo and D. Grimshaw (eds), *Knowledge Intensive Business Services: Organizational Forms and National Institutions*, Cheltenham, UK and Northampton, MA, USA: Edward Elgar, pp. 205–35.

Casper, S., J.R. Hollingsworth and R. Whitley (2005), 'Varieties of capitalism: comparative institutional approaches to economic organization and innovation', in S. Casper and F. v. Waarden (eds), *Innovation and Institutions. A Multidisciplinary Review of the Study of Innovation Systems*, Cheltenham, UK and Northampton, MA, USA: Edward Elgar, pp. 193–228.

Casper, S., M. Lehrer and D. Soskice (1999), 'Can high-technology industries prosper in Germany? Institutional frameworks and the evolution of the German software and biotechnology industries', *Industry and Innovation*, **6** (1), 5–24.

Cooke, P. (2005), 'Regionally asymmetric knowledge capabilities and open innovation. Exploring "Globalisation 2" – a new model of industry organisation', *Research Policy*, **34**, 1128–48.

David, P.A. (1993), 'Path-dependence and predictability in dynamic systems with local network externalities: a paradigm for historical economics', in D. Foray and C. Freeman (eds), *Technology and the Wealth of Nations: The Dynamics of Constructed Advantage*, London: Pinter Publishers, pp. 208–31.

Djankov, S., C. McLiesh, M. Klein, World Bank and International Finance Corporation (eds) (2006), *Doing Business 2007. How to Reform: Comparing Regulation in 175 Economies*, Washington, DC: World Bank Publications.

Dosi, G. (1982), 'Technological paradigms and technological trajectories: a suggested interpretation of the determinants and directions of technical change', *Research Policy*, **11**, 147–62.

Dosi, G., C. Freeman, R.R. Nelson and L. Soete (eds) (1988), *Technical Change and Economic Theory*, London: Pinter.

Edquist, C. (2005), 'Systems of innovation: perspectives and challenges', J. Fagerberg, D.C. Mowery and R.R. Nelson (eds), *The Oxford Handbook of Innovation*, New York: Oxford University Press, pp. 181–208.

EITO (European Information Technology Observatory) and European Economic Interest Grouping (EEIG) (eds) (2006), *European Information Technology Observatory 2006*, Frankfurt: European Information Technology Observatory.

Essletzbichler, J. and D.L. Rigby (2007), 'Exploring evolutionary economic geographies', *Journal of Economic Geography*, **7** (5), 549–71.

Freeman, C. and C. Perez. (1988), 'Structural crises of adjustment, business cycles and investment behaviour', G. Dosi (ed.), *Technical Change and Economic Theory*, London: Pinter Publishers, pp. 38–66.

Friedewald, M., K. Blind and J. Edler (2001), 'Softwareentwicklung in Deutschland. Eine Bestandsaufnahme' [Software development in Germany: a stocktaking], *Informatik Spektrum*, **24**, 81–90.

Garud, R. and P. Karnøe (2001), 'Path creation as a process of mindful deviation', in R. Garud and P. Karnøe (eds), *Path Dependence and Creation*, Mahwah, NJ: Lawrence Erlbaum Associates, pp. 1–38.

Gertler, M. (2003), 'Tacit knowledge and the geography of context, or the undefinable tacitness of being (there)', *Journal of Economic Geography*, **3**, 75–99.

GfK Marktforschung, Fraunhofer-Institut für Experimentelles Software Engineering (IESE) and Fraunhofer-Institut für Systemtechnik und Innovationsforschung (ISI) (2000), *Analyse und Evaluation der Softwareentwicklung in Deutschland: Eine Studie für das Bundesministerium für Bildung und Forschung* [Analysis and Evaluation of Software Development in Germany: A Study for the Federal Ministry of Education and Research], Nürnberg: GfK Marktforschung.

Giddens, A. (1984), *The Constitution of Society: Outline of the Theory of Structuration*, Cambridge: Polity Press.

Grabher, G. (1993), 'The weakness of strong ties: the lock in of regional development in the Ruhr area', in G. Grabher (ed.), *The Embedded Firm – On the Socioeconomics of Industrial Networks*, London & New York: Routledge, pp. 278–95.

Guerrieri, P. (1999), 'Patterns of national specialisation in the global competitive environment', in D. Archibugi, J. Howells and J. Michie (eds), *Innovation Policy in a Global Economy*, Cambridge: Cambridge University Press, pp. 139–59.

Hall, P.A. and D.W. Soskice (eds) (2001), *Varieties of Capitalism: The Institutional Foundations of Comparative Advantage*, Oxford: Oxford University Press.

Heinzl, A. and A. Oberweis (2007), 'Industrialisierung der Softwareentwicklung: Eine Bestandsaufnahme' [Industrialisation of software development: a stocktaking], *Wirtschaftsinformatik*, **3** (49), 157–8.

Holl, F., K. Menzel, P. Morcinek, J.T. Mühlberg, I. Schäfer and H. Schüngel (2006), *Studie zum Innovationsverhalten deutscher Software-Entwicklungsunternehmen* (Entwicklungen in den Informations- und Kommunikationstechnologien, Vol. 2) [Study on the Innovation Behaviour in German Software Development Companies, (Developments in Information and Communication Technologies, Vol. 2)], Berlin: Eigenverlag.

Hornschild, K., S. Kinkel and G. Lay (2003), 'Höhere Wettbewerbsfähigkeit durch produktbegleitende Dienstleistungen: Betreibermodelle im deutschen Maschinenbau' [Higher competitiveness by product-related services: operator models in German mechanical engineering industry], Wochenbericht des DIW Berlin 49/03 (Discussion Paper), DIW Berlin.

Lehrer, M. (2000), 'From factor of production to autonomous industry: the transformation of Germany's software sector', *Vierteljahrshefte zur Wirtschaftsforschung*, **4**, 587–600.

Lehrer, M. (2006), 'Two types of organizational modularity: SAP, ERP product architecture and the German tipping point in the make/buy decision for IT services', in M. Miozzo and D. Grimshaw (eds), *Knowledge Intensive Business Services: Organizational Forms and National Institutions*, Cheltenham, UK and Northampton, MA, USA: Edward Elgar, pp. 187–204.

Lundvall, B. and P. Maskell (2000), 'National states and economic development: from national systems of production and national systems of knowledge creation and learning', in G.L. Clark, M.P. Feldman and M.S. Gertler (eds), *The Oxford Handbook for Economic Geography*, Oxford: Oxford University Press, pp. 353–72.

Lundvall, B., B. Johnson, E.S. Andersen and B. Dalum (2002), 'National systems of production, innovation and competence building', *Research Policy*, **31**, 213–31.

Malerba, F. (2006), 'Innovation and the evolution of industries', *Journal of Evolutionary Economics*, **16**, 3–23.

Malerba, F. and I. Orsenigo (2000), 'Knowledge, innovation activities and industrial evolution', *Industrial Corporate Change*, **9** (2), 289–314.

Malmberg, A. and P. Maskell (2006), 'Perspective. Localized learning revisited', *Growth and Change*, **37** (1), 1–18.

Martin, R.L. (2000), 'Institutional approaches to economic geography', in E.S. Sheppard and T.J. Barnes (eds), *A Companion to Economic Geography*, Oxford & Malden, MA: Blackwell, pp. 77–94.
Martin, R.L. (2006), 'Pfadabhängigkeit und die ökonomische Landschaft' [Path dependency and the economic landscape], in C. Berndt and J. Glückler (eds), *Denkanstöße zu einer anderen Geographie der Ökonomie* [Starting Points Towards a Different Geography of Economy], Bielefeld: Bielefeld University Press, pp. 47–76.
Martin, R.L. and P. Sunley (2006), 'Path dependence and regional economic evolution', *Journal of Economic Geography*, **6**, 395–437.
Maskell, P. and A. Malmberg (2007), 'Myopia, knowledge development and cluster evolution', *Journal of Economic Geography*, **7** (5), 603–18.
Meyer-Krahmer, F. (2001), 'The German innovation system', in P. Larédo and P. Mustar (eds), *Research and Innovation Policies in the New Global Economy. An International Comparative Analysis*, Cheltenham, UK and Northampton, MA, USA: Edward Elgar, 205–52.
Montobbio, F. (2004), 'Sectoral dynamics and structural change: stylized facts and systems of innovation approaches', in F. Malerba (ed.), *Sectoral Systems of Innovation, Concepts, Issues and Analyses of Six Major Sectors in Europe*, Cambridge: Cambridge University Press, pp. 42–69.
Naschold, F. (ed.) (1997), *Ökonomische Leistungsfähigkeit und institutionelle Innovation: Das deutsche Produktions- und Politikregime im globalen Wettbewerb* [Economic Performance and Institutional Innovation: The German Production and Policy Regime in Global Competition], Berlin: Edition Sigma.
Nelson, R.R. (1993), *National Innovation Systems: A Comparative Analysis*, New York: Oxford University Press.
Nelson, R.R. (2002), 'Bringing institutions into evolutionary growth theory', *Journal of Evolutionary Economics*, **12**, 17–28.
Nelson, R.R. and S. Winter (1982), *An Evolutionary Theory of Economics*, Cambridge, MA and London: Belknap Press.
Nooteboom, B. (1999), 'Innovation, learning and industrial organisation', *Cambridge Journal of Economics*, **23**, 127–50.
North, D.C. (1990), '*Institutions, Institutional Change and Economic Performance*, Cambridge: Cambridge University Press.
OECD (2006), *OECD Information Technology Outlook*, Paris: OECD Publishing.
Pavitt, K. (1998), 'The social shaping of the national science base', *Research Policy*, **27** (8), 793–805.
Pelikan, P. (2003), 'Bringing institutions into evolutionary economics: another view with links to changes in physical and social technologies', *Journal of Evolutionary Economics*, **13**, 237–58.
Scott, W.R. (2001), *Institutions and Organizations*, Thousand Oaks, CA: Sage.
Soskice, D. (1999), 'Divergent production regimes: coordinated and uncoordinated market economies in the 1980s and 1990s', in H. Kitschelt, P. Lange, G. Marks and J.D. Stephens (eds), *Continuity and Change in Contemporary Capitalism*, Cambridge: Cambridge University Press, pp. 101–63.
Statistisches Bundesamt (2007), *Strukturerhebung im Dienstleistungsbereich: Datenverarbeitung und Datenbanken 2005* [Structure Survey in the Field of Services: Data Processing and Databases 2005], Wiesbaden: Statistisches Bundesamt.
Statistisches Bundesamt (2008), *Umsatzsteuerstatistik* [Sales Tax Statistics], several volumes, Wiesbaden: Statistisches Bundesamt.
Stille, F. (2003), 'Produktbegleitende Dienstleistungen gewinnen weiter an Bedeutung' [Product-related services gain further significance], Wochenbericht des DIW Berlin 21/03 (Discussion Paper), DIW Berlin.
Storper, M. (1997), *The Regional World: Territorial Development in a Global Economy*, New York: Guilford Press.
Strambach, S. (1997), 'Wissensintensive unternehmensorientierte Dienstleistungen – ihre Bedeutung für die Innovations- und Wettbewerbsfähigkeit Deutschlands' [Knowledge-intensive business services – their significance for Germany's innovativeness and competitiveness], *Vierteljahrshefte zur Wirtschaftsforschung*, **66** (2), 230–42.
Strambach, S. (2002a), 'Germany – knowledge-intensive services in a core industrial economy', in P. Wood (ed.), *Consultancy and Innovation: The Business Service Revolution in Europe*, London: Routledge, pp. 124–51.
Strambach, S. (2002b), 'Organisational innovations in different systems of innovation – the case of Germany and UK', in L. Schätzl and J. Revilla Diez (eds), *Technological Change and Regional Development in Europe*, Heidelberg: Physica Verlag, pp. 250–72.
Strambach, S. and C. Storz (2008), 'Pfadabhängigkeit und Pfadplastizität von Innovationssystemen – die deutsche und japanische Softwareindustrie' [Path dependency and path plasticity of innovation systems – the German and Japanese software industry], *Vierteljahrshefte zur Wirtschaftsforschung*, **77** (2), 142–161.
Streeck, W. (1997), 'German capitalism – does it exist? Can it survive?', *New Political Economy*, **2**, 237–56.
Streeck, W. and K.A. Thelen (eds) (2005), *Beyond Continuity: Institutional Change in Advanced Political Economies*, Oxford: Oxford University Press.

Streich, D. and D. Wahl (eds) (2006), *Moderne Dienstleistungen: Impulse für Innovation, Wachstum und Beschäftigung* [Modern Services: Stimuli for Innovation, Growth and Employment], Frankfurt & New York: Campus Verlag.

Whitley, R. (2000), 'The institutional structuring of innovation strategies: business systems, firm types and patterns of technical change in different market economics', *Organization Studies*, **21** (5), 855–86.

Whitley, R. (2002), 'Developing innovative competences: the role of institutional frameworks', *Industrial and Corporate Change*, **11** (3), 497–528.

Zysman, J. (1994), 'How institutions create historically rooted trajectories of growth', *Industrial and Corporate Change*, **3** (1), 243–83.

Appendix

Table 19A,1 Annex

Industrial sector	1999	2000	2001	2002	2003	2004	2005	2006	Rate of change 1999–2006 (%)
Number of companies (TOTAL)									
72 Computer and related activities	46,544	49,730	53,271	55,297	56,584	59,945	63,046	65,441	40.60
72,1 Hardware consultancy	4601	4835	5076	5170	5295	5549	5791	5894	28.10
72,2 Software companies	19,353	21,482	23,965	25,795	27,633	30,783	33,405	36,719	84.57
72,3 Data processing services	17,154	17,015	17,009	16,669	15,457	14,955	14,754	14,370	−16.23
72,4 Databases	177	215	269	292	285	335	349	380	114.69
72,5 Maintenance and repair	1154	1251	1369	1426	1473	1568	1667	1674	45.06
72,6 Other IT services	4105	4932	5583	5945	6441	6755	7080	7404	80.37
All industrial sectors	2,886,268	2,909,150	2,920,983	2,926,570	2,915,482	2,957,173	3,036,758	3,099,493	7.39
Turnover (in 1000€)									
72 Computer and related activities	34,292,959	39,731,171	47,248,567	45,097,138	43,563,961	45,968,797	48,223,595	50,206,822	46.41
72,1 Hardware consultancy	1,820,627	1,967,834	2,651,317	2,190,117	2,182,996	2,344,464	2,790,291	3,150,133	73.02
72,2 Software companies	13,730,936	17,674,953	22,501,705	20,874,065	19,287,731	21,484,590	23,028,297	24,103,182	75.54
72,3 Data processing services	13,948,695	14,520,388	15,633,183	15,470,490	16,110,350	15,571,002	15,476,349	14,898,618	6.81

72,4 Databases	299,131	352,794	483,451	550,322	252,722	258,532	298,512	578,189	93.29
72,5 Maintenance and repair	708,484	831,390	986,598	619,528	630,287	589,134	626,271	624,502	−11.85
72,6 Other IT services	3,785,086	4,383,813	4,992,314	5,392,616	5,099,875	5,721,077	6,003,876	6,852,196	81.03
All industrial sectors	3,897,312,083	4,152,927,275	4,272,885,186	4,252,562,279	4,248,073,736	4,347,506,204	4,567,396,650	4,930,000,205	26.50
Employees subject to social insurance contribution (TOTAL)									
72 Computer and related activities	263,140	318,094	363,715	372,083	365,285	361,882	370,596	385,758	46.60
72,1 Hardware consultancy	11,513	12,893	13,569	12,901	12,668	13,056	12,646	13,969	21.33
72,2 Software companies	175,691	215,360	253,099	259,149	251,011	246,246	257,901	270,086	53.73
72,3 Data processing services	53,697	64,009	70,202	72,779	74,030	74,827	72,878	71,905	33.91
72,4 Databases	833	1183	1269	1270	1406	1627	1985	4190	403.00
72,5 Maintenance and repair	17,973	19,710	16,947	16,973	16,758	16,193	15,044	14,630	−18.60
72,6 Other IT services	3433	4939	8629	9011	9,12	9933	10,142	10,978	219.78
All industrial sectors	27,402,416	27,749,538	27,741,479	27,506,182	26,889,558	26,458,234	26,112,976	26,285,341	−4.,08

Source: Statistisches Bundsamt (2008), Bundesagentur für Arbeit (2007a, 2007b) several volumes, own calculations,

20 On the notion of co-evolution in economic geography

Eike W. Schamp

Economic geography has recently seen various 'turns' in its way of thinking, adapting and absorbing approaches that have been developed in related disciplines. Evolutionary economics is a case in point that seems to increasingly influence empirical work in economic geography, although the 'evolutionary turn' is still considered 'embryonic' (Martin and Sunley, 2006, p. 396). This state of affairs calls for further reflection on the way in which the 'original' meaning of a concept in another discipline can be adapted to the perspectives of economic geography. As the prefix implies, co-evolution always occurs in association with evolution. While a serious debate on the concept of evolutionary economic geography (Boschma and Frenken, 2006) has already begun, its apparent 'alter ego', co-evolution, needs to be examined more closely as a concept in economic geography. This chapter attempts to shed some light on the possible significance of introducing co-evolution into economic geographical thought. Using an analytical rather than a systemic approach, it argues in favour of a very specific interpretation of co-evolution. In a word, in this chapter co-evolution is not seen as a phenomenon that generally accompanies the evolution of something else but as one that applies to a special case.

If co-evolution is therefore to be understood as a type of alter ego, it is helpful to begin with a brief reflection about evolution and its interweaving with co-evolution. Broadly speaking, there are two different perspectives on evolution; that is, the neo-Schumpeterian perspective which looks at the emergence of 'populations' defined in micro-economic terms, and the systemic perspective where the entire economy is seen as an evolving system. Co-evolution has a totally different meaning in each perspective. The first viewpoint calls for an exact delineation of what is co-evolving with what, whereas in the second perspective co-evolution is 'embedded' in an overall systemic understanding. Looking at the first perspective, section 2 discusses aspects of co-evolution that may be crucial for analyses in economic geography. It should be noted that dynamic processes in technology and innovation that can be analysed using the concept of evolution and co-evolution are socially determined. This makes it possible to introduce 'institution' as a differentiating characteristic of co-evolution.To put it more precisely, this chapter understands co-evolution only as co-evolving populations of well-defined institutions. The following sections discuss in more detail some major concepts of institution and institutional forms that have long been familiar to economic geographers: the firm, the institutional setting, and the region. It will become clear that co-evolution is not seen as an 'ordinary' fact in dynamic processes of spatial and regional development, that co-evolutionary analyses require a thorough reflection on any 'objects' included in the analysis and on mechanisms of co-evolution, and therefore also call for further reflection on methodology.

This chapter is intended to provide a pioneering contribution to the development of a more precisely expressed concept of co-evolution for economic geography. First, various general aspects of the concept of co-evolution are discussed. Second, firms as drivers of technical and regional economic processes, and the 'region' as facilitator of business activity and simultaneously a point of reference for the welfare of the people living there are used as a starting point for reflecting about the usefulness of the concept of co-evolution in economic geography. These have been central subjects in economic geographical research in recent decades.

1. Evolution and co-evolution

There was a time when economic geographers believed it was only possible to explain the present through the influence of long-term 'historical' factors. Thus the present state of places and sectors was a result of the past, the geographical object a one-off event. Later, this 'historical' perspective faded into the background for several decades in favour of spatial and behavioural approaches that placed current factors affecting regional development in the foreground. For a long time geographical innovation research, which took up on Schumpeter's ideas on technological development cycles from the early 1980s on (e.g. Marshall, 1987), favoured a current perspective that, however, was embedded in a temporally dimensioned order. The recently introduced evolutionary approach in economic geography can and should take the temporal dimension of regional economic development more seriously. But is the approach that has been discussed in heterodox economics for some time really suitable for research in economic geography? The growing debate on the theoretical concepts of evolutionary economics and the increasing application of the approach in empirical studies seem to indicate that the answer is simply 'yes'.

Central to the evolutionary economics approach is a new technology; that is, a technical artefact, and the ability to use this. The understanding in social sciences that a technology is socially constructed poses the question of how one must think of this process of social construction. Many neo-Schumpeterian studies in economics and the new evolutionary economic geography are inclined to favour a 'technical' viewpoint, in that they place the development of a new industrial sector as a new technical artefact in the foreground (e.g. Boschma and Wenting, 2007). They thereby associate the emergence of a new technology with the simultaneous emergence of an institutional form that corresponds to the capitalist model of society: the firm. Firms, sectors and – as I will show later – the region can be seen as specific institutional forms that matter in the analysis of current spatial dimensions of capitalism. In this perspective, evolution and institution are integral dimensions of socio-technological processes. Added to that is the problem of different scales – micro, meso and macro – which has not yet been resolved (Dopfer et al., 2004). Geographers in general see the firm as an institutional form at the micro-level, the region as a meso-institution and the national economy as a macro-institution. In accordance with this scheme, the seminal work of Nelson and Winter (1982) is on a micro-analytical scale. However, as Nelson (1995, p. 171) later shows, 'under evolutionary theoretic formulations of the sort that Nelson and Winter (1982) have developed, industrial structure and technical advance interact in complex ways. It is reasonable to say that technology and industrial structure co-evolve'. From this we can conclude, first, that evolution and institution are closely intertwined, and second, that different concepts or rather dimensions of the institution are used. For Nelson refers in one case to the co-evolving industry structure, which can be understood as a particular simultaneously evolving institutional form, and the 'supporting institutions', which can be interpreted as the regulatory systems supporting an evolutionary process, and, therefore, form part of the changing environment of an emergent population. As plausible as this intertwinement of evolution and institution seems to be, it is also highly complex given the versatility of both concepts, 'evolution' and 'institution'.

For a long time, the evolutionary approach favoured a broader systemic understanding of how technologies, sectors and regions emerge. For example, various concepts of complementarity or interrelatedness of emerging technologies and 'technology systems' have been discussed (Carlsson and Stankiewicz, 1991; for a debate see Boschma, 1999). They mainly focused on the emergence of generic technologies and follow-up mechanisms in what became related fields of technology (or sectors). Technology-focused as they were, they often lacked a closer perspective on co-emerging institutional forms (organisations). The existence of sectors ('industries') seemed to be taken for granted, and mechanisms of technological interrelatedness were mainly seen as unidirectional causality. More recently, a different systemic understanding of evolution emerged with the national innovation systems concept (for a recent examination see Lundvall, 2007) and the related regional innovation systems concept (Cooke, 2004). These approaches include societal dimensions such as the educational system, the financial system and so on that foster the development of technical innovations on a national or regional scale, but, here, the region is often taken for granted. Geographers should have a particularly critical view of the regional systems concept, as it seems to favour the understanding of a region as a container where things are happening; however, this chapter argues that the region is a particular and social construct emerging with a new technology. Be that as it may, notions such as interdependence, interrelatedness, interconnectedness or context refer to the complexity of the subject of 'evolution', where everything seems to be connected with everything else unidirectionally and/or in feedback loops. In this systemic perspective on evolution it is difficult to separate *particular* dimensions of co-evolution. In short, the systemic perspective on evolution would require a totally different meaning of the 'co-' in co-evolution, as this chapter suggests.

It is surely easier to adopt a new perspective in a discipline if it takes up subjects that are commonly examined in that discipline. In economic geography this is true of the three institutional levels of social organisation in capitalism referred to above: the firm, the institutional settings of a sector and the region (Schamp, 2005). Based on a neo-Schumpeterian understanding of evolution, these will be used to discuss the related concept of co-evolution in economic geography as a special case in real life. Up to now, economic geographers have used the evolutionary approach in important sub-concepts, such as the 'lock-in' (e.g. Grabher, 1993) or path dependency (most recently Martin and Sunley, 2006) of technical and industrial structural processes within a region. Some publications, however, seem to be far removed from deconstructing and reconstructing these concepts for the purposes of economic geography. They understand them more in a metaphorical way as 'felt' evolution instead of analytical constructs (Martin and Sunley, 2006, p. 396). An evolutionary approach thereby quickly becomes a historical narrative and thus exactly what economic geographers already practised decades ago. This, obviously, is not the intention of this chapter.

It is not self-evident that the theoretical constructs of evolutionary economics can simply be transferred to themes and issues in economic geography (Boschma and Frenken, 2006). Evolutionary economics does not offer a unified body of theory, but consists of a heterogeneous mix of different perspectives. If the evolutionary perspective must be seen as ambiguous, then what is the case with regard to its companion, co-evolution?

The basic element of the evolutionary approach is a population that develops and successfully asserts itself in competition with others by means of mutation, selection and retention, mostly in a changing environment. The definition of a population depends largely on the common practice of individual disciplines. Economic geographers tend to think of a population of firms that form a sector and are based on a specific (new) technology. This idea, that a technology, a sector or region depending on specific technology develops in accordance with certain principles of change, selection and reinforcement in a long-term process, that is, in an 'evolutionary' manner, has long been augmented by the idea that this process is only possible in association with accompanying dimensions. Is this 'co-evolution', or is it 'environment'? In the following, the chapter presents a more restricted understanding of the concept of co-evolution that refers solely to the parallel and reciprocal development of a second population. Malerba (2006, p. 17) recently remarked that 'coevolution goes at the heart of the dynamic analysis of innovation and the evolution of industries', but he also imposed a qualification: 'The challenge for research here is to go to a much finer analysis at both empirical and theoretical levels, and to move from the statement that everything is coevolving with everything else to the identification of what is coevolving with what, how intense is this process and whether indeed there is a bi-direction of causality' (p. 18). Initially it seems plausible that the development of a sector or a region is accompanied, if not determined, by parallel development in quite different areas of social life. However, if this idea is not developed more precisely in theoretical terms then it will revert quite quickly to historical narrative. Figure 20.1 summarises the basic ideas of this chapter in taking up the central issues of an evolutionary approach; that is, emerging varieties in and between populations (heterogeneity), the working of selection mechanisms in favour of only one or some of these variants, and mechanisms supporting the long-term survival of variants (retention). It is claimed that the emergence of a co-evolving (institutional), 'second', population can

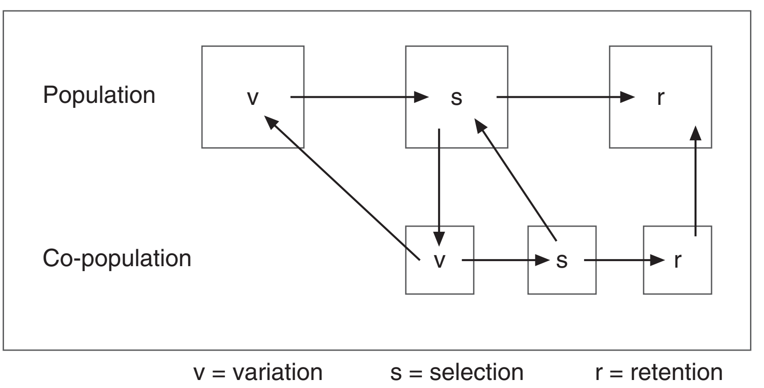


Figure 20.1 A model of two co-evolving populations

function as a selection and retention mechanism for the primary evolving population, among other mechanisms not discussed here. It is assumed, furthermore, that this understanding of co-evolution is of particular interest for geographers if these mechanisms have a 'regional' characteristic such as 'co-location', 'proximity' or 'regional social capital'. In this sense, co-evolution is a 'local' mechanism of self-reinforcement of an evolving (firm) population.

2. Aspects of co-evolution

The evolution of a new technology, a new sector or a new type of region (such as a district or a cluster) always takes place in an environment, which on the one hand is given, but on the other hand also develops continuously. Both form a dynamic, emerging system. The systemic understanding of evolution has an affinity to geographical thought that prefers to address highly complex structures and processes (Martin and Sunley, 2007). It is essential to take this complexity seriously and allow for it in the sharpness of the analytical tools used. What are the specific characteristics of the concept of co-evolution? Malerba's three questions can provide an answer: 'what is coevolving with what, how intense is this process and whether indeed there is a bi-direction of causality'.

Given a multiplicity of systemic relations in which the evolution of a technology, a sector or a type of region is embedded, an exact definition of what is co-evolving is absolutely essential. A precise definition of co-evolution is that two emerging populations exist more or less simultaneously, linked with a mutual causality of development between the evolving and the co-evolving 'population': 'Two evolving populations coevolve if and only if they both have a significant causal impact on each other's ability to persist' (Murmann, 2003, p. 22; see also Malerba, 2006, p. 18). Co-evolution in this sense is technology-specific or sector-specific and thereby dependent on context. Economic geography can learn from empirical studies made by economists and economic historians. First, certain sectors grow in close association with others and could be seen as 'co-evolutionary' (see section 3). Second, research in technology made clear quite early on that a new technology can only be successful with simultaneous transformation in institutional arrangements. Thus sectors and institutional arrangements co-evolve (see section 4). As such institutions are often tied to specific territories, geographical localisation of the co-evolving processes takes place. One example is Murmann's seminal study on the emergence of the dye industry in Germany (2003). Third, economic geographers may see the region as a special form of institution. Storper and Walker (1989, p. 71) introduced the hypothesis that a newly emerging industry creates its own environment. A new sector can go hand in hand with a process of institutionalisation that not only determines the creation of its market (Loasby, 2000) and the regulatory system for its competitiveness, but also includes the creation of a 'region' as a supporting environment (see section 5). Co-evolution in economic geography thereby focusses on the issue of the emerging co-location.

It goes without saying that evolution and co-evolution are time-consuming processes. All the more surprising is the observation that the dimension of time or the temporality of sub-processes has rarely been discussed in this context. There is only a limited debate on the construct of time in economics (Setterfield, 1995) and it is not clear whether the evolutionary approach implies a historical time concept or a Robinsonian 'logical' time concept (see Martin and Sunley, 2006). In evolutionary studies by economic historians, time is an unambiguous historical dimension. The case is similar in economic geography, which can take up from its old position in its 'historical' research orientation. The historical viewpoint, however, often suffers from the weakness that it examines complex interrelations less in an analytical manner, and instead adheres to a descriptive and phenomenological viewpoint. Therefore, a co-evolutionary approach would require a more detailed analysis of the varying temporality of the co-evolving populations. With the concept of 'selective historical dependence', Setterfield (1995, p. 22f.) calls for a selection of the temporal context but at the same time warns of the dangers of 'historical determinism'. The extent of the temporal dimension is determined by the object of research and its scale: long-term processes, for example, are involved in the co-evolution of firm populations, short-term processes in the development of routines and practices in firms. Co-evolution in the narrower sense can only exist if processes take place more or less simultaneously and interdependently; it is neither mere 'contingent' nor 'context' nor an accompanying or subsequent event. This does not mean, however, that co-evolutionary processes must therefore always take place at the same pace and over the same lifespan. Martin and Sunley (2006) have rightly pointed out the positive lock-ins that are absolutely essential in co-evolution, which make a sectoral path or a regional path possible. Institutional co-evolution brings about the breaking of traditional institutions such as traditional sectors, rules, or regions in favour of enabling new technologies, sectors, rules and regions. Later lock-ins run the risk of disadvantageous contraction. This shifting of characteristics poses questions with regard to historically specific events: How does co-evolution break down and what does this mean for the sectors and regions affected (see Schamp, 2005; also see Hassink, Chapter 21 in this volume)?

Co-evolution in the sense proposed here therefore only exists under three conditions: (1) it must be possible to differentiate between two populations that (2) more or less simultaneously (3) have a reciprocal causality. However, other disciplines have different, sometimes 'softer' concepts of co-evolution. For example, Dopfer and Potts (2004) refer in their evolutionary realism approach to the 'bi-modality' and synchronicity of a population of rules and of a population of supporting agents in order to account for evolution (not co-evolution!). In contrast, in organisation science, business models resulting from a specific searching behaviour among new industries are also seen as co-evolving (e.g. Huygens et al., 2001). Where are the boundaries of the concept of co-evolution? This chapter would argue here for a stringent interpretation of the concept. It is in this sense that the following sections examine the concept of co-evolution as applicable to traditional subjects of empirical research in economic geography.

3. The co-evolution of two firm populations

For some classical economic geographers, it was self-evident that certain industrial sectors developed jointly in the eighteenth and nineteenth centuries, for example coal and steel or textiles/shoes and mechanical engineering. Geographers have discussed this issue for the twentieth century with a less definite outcome, for example plastic injection moulding and toolmaking or mail order business and related services (see Schulz et al., 2006) or internet providers and venture capital firms (see Zook, 2002). When economic geographers or economic historians addressed the emergence of linked industrial sectors, this usually involved little more than a historical narrative. There was no general analysis of the mechanisms of evolutionary interaction. Economic geographers were only interested in the description of shared histories when both sectors were co-localised in so-called local production systems. Old industrialised areas were the preferred subject of research in economic geography, such as the Ruhr region or textile producing regions.

Insofar as current research in economic geography is concerned with different sectors in local production systems, it is mainly interested in the mechanisms by which they function. Therefore supply linkages between vertically linked sectors are usually examined, followed by a description of linkages of knowledge. This research is in itself useful and may also produce valuable information for subsequent evolutionary analyses. Nevertheless, it must be remarked that the analysis of the processes of co-evolution of interlinked sectors in local production is just beginning.

Recent studies have first tried to address the issue of how a population of firms evolves in the 'core sector' of a local production system or 'cluster', a concept that has become very popular in the last few years. Beginning with earlier studies, for example on the emergence of Silicon Valley (Saxenian, 1985) or the so-called Cambridge phenomenon (Garnsey and Heffernan, 2007; Segal Quince & Partners, 1985), and continuing, for example, with the analysis of the development of a cluster in the German packaging industry (Mossig, 2005), responsibility for the development of clusters was attributed to spin-offs and spin-outs. That spin-offs have greater chances of survival compared to new foundations and that clusters can only emerge in this way have recently been demonstrated by Boschma and Wenting (2007). They explain these findings with the selective adoption of successful routines from incubator firms as well as in firms using related technologies (e.g. automobile from bicycle production). Therefore, two processes contribute to the evolution of a sector: first selective replication of successful routines, and second variation of these routines leading to a diversification in the sector. In these studies co-location remained somewhat self-evident because of an increased likelihood that spin-offs deliberately choose geographical proximity to their incubator instead of looking for 'arbitrary' locations elsewhere. There is no lack of plausible explanations from traditional locational studies of small and medium-size enterprises (SMEs). A low level of information on alternative locations and attachment to the founder's existing lifeworld are two such explanations (Schamp, 2000) used to account for the emergence of co-localisation within a sector. But what explains the co-localisation of co-evolving sectors?

The reason for the co-evolution of supporting sectors in a cluster seemed to be just as self-evident as co-location. Economic geographers, historians and economists have all failed to provide an explanation based on evolutionary theory to show why supporting sectors emerge as new populations of firms. The evolution of a new sector always seems to be associated initially with a marked vertical integration of its firms. This can be accounted for by the fact that incumbent sectors have difficulty in meeting the technical requirements of the emerging sector. In consequence, the question arises as to when and why the integrated firms of an emerging sector start to buy products instead of producing them themselves. This is the classical question addressed by transaction cost theory. In its more or less static form – given that transaction costs determine whether market, network or hierarchy is the best way to coordinate economic activities – it cannot contribute an answer, as the factors that determine the size of transaction costs remain exogenous. Jacobides and Winter (2005) argue in favour of a 'systemic' perspective on the correlation between firm-internal capabilities and endogenously altered transaction costs to better understand the dynamism of co-evolutionary processes. They suggest linking learning-associated changes in the capabilities of a firm with changes in transaction costs in such a way that the altered scope of a firm can be explained. A reduction of the scope of the firm implies the de-integration of the vertical chain and the development a new supplier sector. However, for Jacobides and Winter this is not an irreversible process. Using the example of two sectors they show how, first, specialisation through outsourcing, and second, renewed intensification of scope through insourcing can take place over a medium-term period of about ten years.

The starting point for their considerations is the given heterogeneity of firms in a sector with regard to production costs and the vertical extent of their capabilities. Insofar as the more specialised firms have a cost advantage, the market will likewise force the others towards specialisation, and vice versa; this is the 'classical' selection of evolutionary economics. Which process dominates is determined by the respective sector, but is also subject to the learning processes of each individual firm. If firms succeed in changing their environment with regard to transaction costs then a process of variation of the scope of the firm and a selection of what is then the 'best' path begins. If the best path is vertical de-integration, this can lead to a cumulative process of knowledge accumulation through specialisation, which in turn drives the creation of a new supply sector.

In this approach to co-evolution oriented toward change in institutional conditions, a new but reversible sector emerges, dependent on the firm-internal capabilities, firm-internal learning and changing transaction costs of a 'core sector'. This, however, raises new questions. According to Jacobides and Winter (2005), a new technology, quartz technology, was responsible for the reintegration of the watchmaking industry, particularly in the Swatch company. But is the watchmaking industry of 2000 the same as that of the 1980s? What justifies speaking of reversibility and not of fundamental reinvention of the (quartz) watchmaking industry? The theory of the endogenisation of changes in transaction costs through learning processes in the firm and selection processes in the sector is a useful insight, but it cannot ignore the fact that exogenous processes continue to have considerable influence on transaction costs. Current trends towards outsourcing and offshoring of business processes may be associated with painful learning processes for individual firms, but are subject to further conditions that are now summarised under the heading of 'globalisation'.

Jacobides and Winter (2005) presented a conceptual framework for the co-evolution of sectors but made no reference to the necessity of co-location. If economic geographers are primarily interested in co-location as has been the case up to now, further arguments are necessary to explain why the mechanisms of co-evolution should lead to the spatial proximity of the affected sectors. The answer is sought in specific spatially tied institutions that facilitate or enable co-evolution in spatial proximity. According to Brenner (2004, p. 23ff.) one should, however, differentiate more precisely between three types of self-reinforcing process in clusters: direct cooperation processes between firms, the 'indirect' appearance of local externalities (see Schamp, 2002), and the emergence of associated supplier firms for goods and services. It has remained unclear to what extent local 'symbiosis with other industries' can actually be proven (Brenner, 2004, p. 46). In his attempt to understand the evolution of industrial clusters in a general sense by developing a model, the co-evolution of associated firms had largely to be excluded.

The co-evolution of populations of firms as a long-term process is very plausible.

Nevertheless, its significance for research in evolutionary economic geography it not yet clear. Co-evolution associated with co-location is apparently a special case that obviously requires a particular explanation. But this very case is of particular interest to economic geography, which is concerned with the development of regions. A hitherto neglected aspect is the problem of the survival of clusters and its significance for regional development. It appears that those clusters can survive longer which have associated, formerly co-evolving sectors, and that the 'death' of a cluster is more likely to offer potential for a new beginning in a region when associated sectors are present (e.g. see Schamp, 2005 for a cluster in shoe production). Evolutionary studies have an answer for such mechanisms if they underline the importance of previous technical experience in (technically) 'related firms and sectors' (Boschma and Wenting, 2007).

Equating the co-evolution of a sector with its co-location is, however, too narrow a viewpoint for economic geography. Why should co-evolution be automatically associated with co-location? The problem thus becomes more complex: co-evolving sectors with *differing* locational logic can apparently only be understood if the evolution of the associated sector in *its own* location is known. Such an understanding of co-evolving sectors would require a broader geographical approach including different locations ('locationally systemic' approach). In the age of globalisation this would surely be an interesting perspective.

For co-evolution of two firm populations to be studied in economic geography, two reservations should be considered. First, bearing bi-causality in mind, it is still necessary to place the processes of reciprocal reinforcement in co-evolution at the foreground. For there are many sectors whose supply of goods and services is important for the evolving sector, but whose own evolution is dependent on that of the new sector only to a very limited extent or not at all. Second, this implies that co-evolution between sectors is possibly a fairly rare event dependent on special technical conditions. Therefore one must assume a limited number of co-evolving sectors around a core sector.

4. The co-evolution of populations of firms and institutional arrangements

As already mentioned, institutions are at least as versatile a construct as evolution and co-evolution. Pelikan (2003, p. 238) expressed this very clearly in response to Nelson (2002): 'should institutions be defined as constraints, comparable to "the rules of the game", or as routines, meaning "specific ways of playing the game"?' Should the first definition, formulated by North (1990) and which refers to the meso and macro levels, or the second more micro-analytical definition discussed by Nelson (2002) be used? Pelikan gives detailed reasons for his preference for the first definition, which corresponds with the debate in economic geography on the 'institutional turn' (Hayter, 2004; Martin, 2000). I therefore begin by using this concept of institution in the following in order to analyse the way in which co-evolution takes place in association with a technology. However, both definitions may not be as far apart as has been assumed. Could it be that new 'specific ways of playing the game' may also sometimes change the 'rules of the game'? Does this indicate a possibility for identifying the evolution of institutions? I discuss this at the end of this section.

The following does not include a discussion of the fact that the problem of the connection between institutions and economic change can be traced back to Marx, if not even earlier, and then to Veblen (e.g. Nelson, 2002, p. 19). A line could certainly be drawn from Marx via Schumpeter to the neo-Schumpeterians. So this is an old problem of economics as a social science. In order to explain the co-evolution of a specific technology that expresses itself in a new sector, with an institution in the sense of regulatory systems that then also must find an organisational form of their own, one must be more specific. Remember Malerba (2006): he calls for proof of 'bi-causality'; that is, a reciprocal dependence of the evolving sector and specific associated institutions. If we are to take the concept of co-evolution seriously, then evolution and institution are always specific in the context of a specific reciprocal dependency. One viewpoint that readily suggests itself to economists focuses on the connection between evolution (of a technology/sector) and the construction of a market for the new technology (Loasby, 2000). This takes the output aspect of the technology process into account. Bearing in mind the narrow definition of the concept of co-evolution discussed at the beginning of this chapter, I will not examine this in greater detail. Institutional arrangements that directly influence the development of a technology at the input level seem more interesting for research in economic geography, because they are more strongly connected to the traditional perspectives of the discipline. Here, the spatial organisation of production matters more than consumption.

The important characteristic of institutions as 'rules of the game' is their function in regulating the selection process of a technology or sector. If institutions develop simultaneously with the technology they ultimately contribute to the reinforcement of a technology path or a sector. Using the example of competition between the British, American and German dye industry in the nineteenth and early twentieth centuries, Murmann (2003) has shown how a competitive advantage (Nelson, 1995) emerges in the evolutionary process. In his detailed study, Murmann analyses the co-evolution of two different national institutions – research and training systems and patent practices – with the evolution of the German dye industry. In this way he shows why the German dye industry became the worldwide leader in the sector, although it started later than its British counterpart. This study has a number of interesting aspects that are useful for the debate on co-evolution in economic geography. First, only those institutions are taken into account for which a 'bi-directional causality' between technology and institution is provable (Murmann, 2003, p. 23). Murmann demonstrates, for example, the influence of the dye industry on the development of research at German universities and the influence of university lecturers on the dye industry. These are not necessarily all the institutions that develop simultaneously with the technology. Murmann, for example, also points out the accompanying learning processes in the financial sector and the growing strength of the supplier industry in Germany (p. 196), without however analysing these in greater detail. The co-evolutionary perspective therefore requires that specific institutions must be chosen for the analysis of the emergence of a specific technology. Second, he extends the co-evolutionary perspective to 'multiple things [that are] jointly evolving'. Thus the evolution of a technology is seen in co-evolution with a bundle of institutions; in this case two co-evolving institutions are identified and analysed (p. 21). Third, while pre-existing institutional arrangements can make the evolution of a technology/sector easier, the simultaneity of evolutionary processes in the sector and the specific institution is of the greatest significance. This is an eminently social process, as can be seen in the early formation of associations in Germany that influence the 'social construction' of the institutions. Fourth, this has implications for further aspects of co-evolution. For co-evolution is not deterministic, but is characterised by trial and error, which means that the timely recognition and elimination of mistakes has major significance for the explanation of the resulting co-evolution. Or, as Murmann (2003, p. 222) puts it: 'no failures, no evolutionary adaption through selection'. The timing of individual events that ultimately have an influence on shared evolution is important.

Many aspects brought up by Murmann are already being discussed in economic geography. That firms are not a product of their environment, but can shape it in order to develop, has been discussed by Storper and Walker (1989), for example. Collective activity is essential, as an individual firm can hardly shape its environment on its own, an issue that has been discussed in the debate on the emergence of industrial districts (Schmitz, 1995). In the broader sense we can conclude from this that the co-evolution of a technology/sector and institutions can lead to positive external effects that favour and establish the development path of a sector in competition with other sectors in the territories in which the institutions apply. Under the assumption that most institutional arrangements are tied to a specific territory (whether at a national level as in Murmann's study, or at a regional one as in the debate on industrial districts) the co-evolution of sector and institution also explains the existence of a differentiated geography (Hayter, 2004). Or to put it the other way around: co-evolution of a new sector and its reinforcing institutions may only take place where regional or national societies are prepared to overcome traditional rules and habits and are sufficiently ingenious to create new ones. Current research in economic geography can take up on two aspects of this. First, a comprehensive debate has developed on the significance of national and regional innovation systems (see Bathelt, 2003; Cooke, 2004; Iammarino, 2005). As mentioned before, however, these often see the dynamics of technology, sector and institution as linked inseparably, which does not pave the way for a specific analysis of co-evolution. Second, the concept of the (regional) lock-in of sectors (and their technologies) has been discussed in economic geography for quite some time. On the one hand, this debate is concerned with a positive aspect of lock-in in path-dependent processes (Martin and Sunley, 2006), which is emphasised here with the co-evolution of technology and institutions. On the other hand, when and why the simultaneity and reciprocality of co-evolving units becomes dysfunctional in the form of negative lock-in often remains empirically unclear. The same is true for the question of when it collapses, thereby inducing the end of both the technology/sector and the supporting institutions. Geographers have been particularly interested in the negative aspects of lock-in in institutional rigidities accompanying ageing technologies/sectors, and hence the non-simultaneity of technical and institutional development in the late phases of a technology/sector (e.g. Grabher, 1993; Hassink and Shin, 2005; Schamp, 2005).

In many respects the debate on the co-evolution of institutions is itself still at a stage prior to that of final selection and retention. It is still a new debate and there are different hypotheses that differ in the scope of the institutional concept, the temporality of the processes to be analysed and the exactness of the links to the evolving technology. One level of differentiation could be the period of time that is analysed, and thus the choice of methodology. The co-evolution of the German dye industry covers a period of about 70 years (Murmann, 2003), for which the combined action of technology, sector and a few institutions has been described in great detail. For an analysis of the very long-term processes of the three industrial revolutions, Von Tunzelmann (2003) suggests a very broad network approach referring to 'governance' as an 'organizing collective action'. This broad network approach accepts that 'there is no necessary reason for these [i.e. path of technology creation and governance evolution] to coincide either in space or time' (von Tunzelmann, 2003, p. 379). This approach is close to the understanding of a national or regional innovation system often adopted in geography, which is not concerned with a specific technology and a specific 'bi-causality'.

Is the co-evolutionary approach therefore only useful for a long-term perspective in economic geography, a perspective that is not central to economic geography today? I do not think so. Let us return to the example of the dye industry. Murmann has shown how in a fairly short period the emergence of a new technology (product) became possible through new interlinking with research institutions, as part of long-term co-evolution. It can be concluded from this that current processes in co-evolution can also be analysed even if it is not possible to deduce a long-term development path from them. Obviously the latter can only ever be identified after the fact. In this context it should be possible to see the Pelikan–Nelson controversy in a new light. Nelson (2002) sees in the routines the concept that unites evolution and institution. According to him, routines are not merely 'physical' technologies, but also 'standardised patterns of human transactions' or a 'social technology'. One of Pelikan's arguments for not following Nelson in his definition of institutions refers to the issue of the extent to which an institution allows agents to behave in different ways: a routine is (too) narrow a guideline, while a rule, as defined by North, is a limiting framework for several behavioural options (Pelikan, 2003, p. 244). While I sympathise with Pelikan and prefer North's view of institutions, it still seems possible that changes in routines may not only be the result but also the cause of changes in institutions (as rules). Consider for example new routines in labour divided on the basis of information technology, which, when they become standards, are likely to allow new sectors and organisations such as call centres, consultancy firms and so on to develop in new locations. Current offshoring processes could thus be interpreted and analysed from the point of view of economic geography as co-evolutionary processes.

In this way a debate can develop in economic geography that does not see the evolution of a technology/sector as the result of co-evolving institutions external to the companies, but increasingly sees this as the product of evolution within the companies. This can include such dissimilar phenomena as the acquisition of competences through outsourcing to a newly emerging sector, the learning of new technologies through company-internal competence in 'related' technologies or the establishment of a new business form such as the co-operative, an impressive case study of which has recently been provided by Frenken and van der Steege (2006) using the example of the Dutch dairy industry.

5. The co-evolution of regions and populations of firms – a special case for geographers

The regional perspective remains a central focus of geography. This construct is therefore a test for the usefulness of the adoption of the evolutionary approach and the concept of co-evolution in economic geography. As processes in time – not in space – are the central theme of evolutionary theory, a static view such as in the discussion on different types and concepts of what a region might be, which geographers sometimes pursue in exhaustive detail (Allen et al., 1998; Paasi, 2004), is not appropriate here. Rather the question arises as to how a region as a specific territorial unit emerges in combination with an emerging new firm population based on a new technology. It is in this – restricted – sense that I referred to the region as a specific (though admittedly highly complex) form of institution in the opening section of this chapter. This is not to say that co-evolutionary processes may operate at various spatial scales. The focus here is on the necessity of co-emerging important local institutions if co-evolution of firm population and region are assumed to exist. If this view is accepted, then the co-evolution of sector (technology) and region in the context of the bi-causality called for by Murmann (2003) or Malerba (2006) requires that the region constitutes itself in close association with the emerging new firm population and technology, and that this technology can only become established in competition with other technologies through the co-emerging local institutions. This has two implications: first, the region is then understood as a specific social construct, which constitutes itself through networks, shared knowledge and shared ideas and so on via and in association with the relevant technology or the firms representing this technology; second, this region is linked exclusively to the technology in question, or its firms, at both the temporal and the spatial level. The region can therefore be temporary and variable in its extent. The view of co-evolution as an institutionalisation process of the region calls for a detailed analysis of the interaction of technology and institutionalisation, as suggested in section 4 and exemplified by Murmann's analysis of the German dye industry.

This view of the co-evolution of firms (technology) and regions corresponds most closely to the concepts of the district and cluster as they are seen in economic geography. The effects of the co-location of firms using a particular technology on the growth of the firms have been addressed by several authors under the heading of processes of self-reinforcement and self-organisation (e.g. Brenner, 2004), collective order (Scott and Paul, 1990), collective efficiency (Schmitz, 1995) or in more general terms creating local increasing returns (Schamp, 2002). This alone does not appear sufficient for a co-evolutionary approach, as frequently only a uni-causality is examined, for example how local immobile factors such as untraded interdependencies (Storper, 1997) encourage the development and persistence of a cluster. Co-evolution in the sense referred to here would only exist if the particular regional institutions that reinforce the evolution of the technology came into being with the development of the technological cluster, and, as a result, a new 'region' came into existence. Likewise, the co-evolution of a sector or technology and the region would take place if a specific local form of path dependence or 'place dependence' in technological development could be proven to exist (Martin and Sunley, 2006, p. 409).

Various attempts to understand the geographical region as an institution point the way. Hayter (2004), for example, emphasises Holmén's (1995) view of the region as a 'meeting place' of interests, as an arena in which conflicts over resources, including those for the development of firms in a specific new technology, are played out. This is certainly a starting point for the examination of ways in which technology and region may co-evolve. Nevertheless, this alone would not be sufficient, as the formation of regions as an institutionalisation process must be thought of multidimensionally. Co-evolution only begins to take place with the development of new institutions that are tied to a region and simultaneously encourage the development of the specific firms (technology).

In this very narrow and specific sense of the co-evolution of firms and regions there is still a fundamental lack of empirical data. Clusters can also develop in other ways, such as through the effect of static-urban agglomeration advantages, as discussed in the new geographical economics after Krugman. The positive externalities referred to in this context developed 'unintentionally' through individual actions. Co-evolution in the sense discussed in this chapter, however, refers per se to a dynamic cognitive process of reciprocal influence, in which technology and purposeful actions are linked. The resulting institutions are understood to be consciously aimed for (at least by some) and deliberately utilised. The analysis of co-evolution and regionalisation therefore calls for a similar socio-political perspective, as for instance used by Murmann (2003) for the co-evolution of the dye industry and national institutions.

Thus the construct of regional path dependency currently debated in economic geography (Martin and Sunley, 2006) may not be a sign of co-evolution if it is seen in the uni-directional sense of the past influencing the present. I would like to expand on the problems of distinguishing between path dependency and co-evolution using the example of Frankfurt's growth as a financial centre (Grote, 2003, 2004). Although a satisfying answer cannot be found to the question of whether co-evolution was a factor in the institutionalisation of the financial centre, a glance at it is helpful for our debate. Following the division of Germany in 1945, Frankfurt gradually replaced Berlin as the leading financial centre and attained a unique position among Germany's financial centres (see Schamp, 1999). Today, Frankfurt as a financial centre is subject to European competition. Grote has identified different phases in the specific agglomeration advantages for information processing in the financial centre that make the location attractive to German and foreign banks. They grow and decline, often as a result of technical innovations and regulatory changes at a national or international level. In this way he identifies a long-term path dependency in the emergence of the financial centre as a whole, while also demonstrating the changes in the reinforcing mechanisms in the different phases; that is parts of the development path. In brief, for such a complicated object as a (local) financial centre, 'path dependency' and 'co-evolution' are not synonymous. Rather, it is necessary to examine specific aspects of the spheres of activity of a financial centre and of the temporal dimension in order to analyse the emergence of a business segment in the financial sector in association with emerging, locally tied rules. In the case of Frankfurt this could be the specific business segment of mergers and acquisitions of and for German firms (Lo, 2003).

Lo examined different forms of a mechanism very familiar in current institutional economic geography, which are differently suited to demonstrate the co-evolution of a business segment and its own local institutional arrangements. A fairly calculated, knowledge-based trust obtains between the service providers involved in the mergers and acquisitions process, a trust that can be immediately reversible in individual cases, while an identification-based, long-term, personal trust is absent. Nevertheless this short-term, calculating trust has a long-term effect on the emergence and persistence of the mergers and acquisitions sector in Frankfurt. Two conclusions can be drawn: path dependency and cluster may appear together in space, but not necessarily simultaneously in time. Path dependency and cluster should not inevitably be understood as processes in the co-evolution of cluster-specific local institutions. It seems to me that we are far from really understanding co-evolution between technology-based clusters and regionally based institutions, and I believe that this kind of co-evolution of firms (technology) and regionalisation is a special case in geography whose existence must be specifically proven.

6. Conclusion

The idea that technical evolution is socially constructed implies that technology and its institutional embeddedness are inseparable. It is not clear how this demonstrates co-evolution in the narrower sense of 'bi-causality' that we have used here. For all historical experience has shown that a newly emerging technology is always implemented in new institutional forms. In this chapter I differentiated between three institutional forms, the firm (or the regulatory systems prevalent in the firm that Nelson and Winter, 1982, call routines), the 'social' regulatory systems affecting firms (which often exist at a national or even international level), and the region as a specifically institutional construct that is a major focus of geographical interest. So, this is all about evolution and co-evolving institutions. The justification for analysing co-evolution and co-evolving institutional 'populations' was perceived as lying in its capacity to explain mechanisms of selection and retention among competing evolving populations of technology and firms. Murmann's study of the competing US, British and German dye industries in the nineteenth century presented an impressive example of this. Theoretical discussion and empirical analysis of the co-evolution of technology and the first two types of institution are only just beginning, as the preceding sections have made clear. This is also true for the third type of institution. The issue of the possible co-evolution of technology and institutions is a little more complex for geographers than for other social scientists, as the former see the region as a further dimension. If, and only if, co-evolution of another firm population (section 3) and/or co-evolution of institutional arrangements do emerge locally, that is, if 'place' matters in a very restricted sense, can geographers speak of co-evolution of regions.

What consequences could this have for future research in economic geography? I am far from being able to develop a full research programme from this chapter. In order to start with research on co-evolution in economic geography, Malerba's advice should be taken seriously: 'go to a much finer analysis'. This would require:

- Careful selection of cases where the concept of co-evolution may apply – not 'everything is co-evolving with everything else' – that is, a selection of those cases in which co-evolution does not simply refer to more or less simultaneous processes of technology being conditional on institutions, but where reciprocal ('bi-causal') processes occur.
- Precise pre-selection of units of co-evolving populations for empirical studies.
- Further theoretical refinement of the concepts of institutions and institutional change – maybe new institutions such as dynamic transaction costs or 'old' institutions.
- For geographers in particular, a focus on why co-location of two evolving populations matters; that is, selection of those cases where the conditionality of reciprocal development requires an institutional region.
- Also, careful reflection on methodology. Following Staber (2007) who recently called for a 'contextualised' view of social capital in regional clusters, one may suggest dense description as a means to clearly identify and separate 'environment' from 'co-evolving population' in empirical studies, to focus on processes and events, thus emphasising dynamism in historical time, and to be precise in analysing the mechanisms of reciprocal self-reinforcement or retention of the two or more populations considered.

This contribution was written with a fundamental sympathy for the perception of economic geography as an evolutionary science. The concept of co-evolution seems to me to describe a particularly challenging programme for economic geography, which can lead to a better understanding of the dynamics of technical and spatial-economic development. This concept is laden with pre-conditions. It was the aim of this chapter to make this clear.

Acknowledgement
I wish to express my thanks to Johannes Glückler for comments on an earlier draft of the chapter, and Ron Boschma on a later version. I am indebted to Edel Sheridan-Quantz for professionally translating this text, addressing quite a complex matter.

References

Allen, J., D. Massey and A. Cochrane (1998), *Rethinking the Region*, London: Routledge.
Bathelt, H. (2003), 'Geographies of production: growth regimes in spatial perspective I – innovation, institutions and social systems', *Progress in Human Geography*, **26** (6), 763–78.
Boschma, R. (1999), 'The rise of clusters of innovative industries in Belgium during the industrial epoch', *Research Policy*, **28**, 853–71.
Boschma, R.A. and K. Frenken (2006), 'Why is economic geography not an evolutionary science? Towards an evolutionary economic geography', *Journal of Economic Geography*, **6** (3), 273–302.
Boschma, R.A. and R. Wenting (2007), 'The spatial evolution of the British automobile industry. Does location matter?', *Industrial and Corporate Change*, **16**, 213–38.
Brenner, T. (2004), *Local Industrial Clusters. Existence, Emergence and Evolution*, London: Routledge.
Carlsson, B. and R. Stankiewicz (1991), 'On the nature, function, and composition of technological systems', *Journal of Evolutionary Economics*, **1**, 93–118.
Cooke, P. (2004), 'Introduction: regional innovation systems – an evolutionary approach', in P. Cooke, M. Heidenreich and H.J. Brazcyk (eds), *Regional Innovation Systems: The Role of Governance in a Globalized World*, 2nd edn, London: Routledge, pp. 1–18.
Dopfer, K., Foster, J. and J. Potts (2004), 'Micro–meso–macro', *Journal of Evolutionary Economics*, **14**, 263–79.
Dopfer, K. and J. Potts (2004), 'Evolutionary realism: a new ontology for economics', *Journal of Economic Methodology*, **11** (2), 195–212.
Frenken, K. and G. van der Steege (2006), 'The evolution of the Dutch dairy industry and the rise of cooperatives: combining transaction cost and evolutionary approaches', *Papers in Evolutionary Economic Geography*, No. 06.08, available as pdf from: http://econ.geog.uu.nl
Garnsey, E. and P. Heffernan (2007), 'The Cambridge high-tech cluster: an evolutionary perspective', in K. Frenken (ed.), *Applied Evolutionary Economics and Economic Geography*, Cheltenham, UK and Northampton, MA, USA: Edward Elgar, pp. 27–47.
Grabher, G. (1993), 'The weakness of strong ties: the lock-in of regional development in the Ruhr area', in G. Grabher (ed.), *The Embedded Firm: On the Socioeconomics of Industrial Networks*, London: Routledge, pp. 255–77.
Grote, M. (2003), 'Die Evolution des Finanzplatzes Frankfurt' [The evolution of the financial place Frankfurt], *Geographische Zeitschrift*, **91** (3+4), 200–17.
Grote, M. (2004), *Die Entwicklung des Finanzplatzes Frankfurt. Eine evolutionsökonomische Untersuchung* [The Development of the Financial Place Frankfurt. An Evolutionary Analysis], Berlin: Duncker & Humblot.
Hassink, R. and D.-H. Shin (2005), 'Guest editorial: The restructuring of old industrialised areas in Europe and Asia', *Environment and Planning A*, **37**, 571–80.
Hayter, R. (2004), 'Economic geography as dissenting institutionalism: the embeddedness, evolution and differentiation of regions', *Geografiska Annaler*, **86B**, 95–115.
Holmén, H. (1995), 'What's new and what's regional in the "New Regional Geography"?', *Geografiska Annaler*, **77B**, 47–63.
Huygens, M., Ch. Baden-Fuller, F. Van den Bosch, H. Volberda (2001), 'Co-evolution of firm capabilities and industry competition: investigating the music industry', 1877–1997. *Organization Studies* **22**, 971–1011.
Iammarino, S. (2005), 'An evolutionary integrated view of regional systems of innovation: concepts, measures and historical perspectives', *European Planning Studies*, **13**, 497–519.

Jacobides, M.G. and S.G. Winter (2005), 'The co-evolution of capabilities and transaction costs: explaining the institutional structure of production', *Strategic Management Journal*, **26**, 395–413.

Lo, V. (2003), *Wissensbasierte Netzwerke im Finanzsektor. Das Beispiel des Merger & Acquisitions-Geschäfts*, [Knowledge Based Networks in the Financial Sector. The Case of Mergers & Acquisitions], Wiesbaden: Deutscher Universitätsverlag.

Loasby, B.J. (2000), 'Market institutions and economic evolution', *Journal of Evolutionary Economics*, **10**, 297–309.

Lundvall, B.-A. (2007), 'National innovation systems – analytical concept and development tool', *Industry and Innovation*, **14** (1), 95–119.

Malerba, F. (2006), 'Innovation and the evolution of industries', *Journal of Evolutionary Economics*, **16**, 3–23.

Marshall, M. (1987), *Long Waves of Regional Development*, Basingstoke: Macmillan.

Martin, R. (2000), 'Institutional approaches in economic geography', in E. Shepherd and T.J. Barnes (eds), *A Companion to Economic Geography*, Oxford: Blackwell, pp. 77–94.

Martin, R. and P. Sunley (2006), 'Path dependence and regional economic evolution', *Journal of Economic Geography*, **6**, 395–437.

Martin, R. and P. Sunley (2007), 'Complexity thinking and evolutionary economic geography', *Journal of Economic Geography*, **7**, 573–601.

Mossig, I. (2005), 'The evolution of regional packaging machinery clusters in Germany', in A. Lagendijk and P. Oinas (eds), *Proximity, Distance and Diversity. Issues on Economic Interaction and Local Development*, Aldershot: Ashgate, pp. 155–67.

Murmann, J.P. (2003), *Knowledge and Competitive Advantage. The Coevolution of Firms, Technology, and National Institutions*, Cambridge: Cambridge University Press.

Nelson, R.R. (1995), 'Co-evolution of industry structure, technology and supporting institutions, and the making of comparative advantage', *International Journal of Economics of Business*, **2** (2), 171–84.

Nelson, R.R. (2002), 'Bringing institutions into evolutionary growth theory', *Journal of Evolutionary Economics*, **12**, 17–28.

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

North, D.C. (1990), *Institutions, Institutional Change and Economic Performance*. Cambridge, UK: Cambridge University Press.

Paasi, A. (2004), 'Place and region: looking through the prism of scale', *Progress in Human Geography*, **28** (4), 536–46.

Pelikan, P. (2003), 'Bringing institutions into evolutionary economics: another view with links to changes in physical and social technologies', *Journal of Evolutionary Economics*, **13**, 237–58.

Saxenian, A. (1985), 'The genesis of Silicon Valley', in P. Hall and A. Markusen (eds), *Silicon Landscapes*, Boston: Allen & Unwin, pp. 20–34.

Schamp, E.W. (1999), 'The system of German financial centres at the crossroads: from national to European scale', in E. Wever (ed.), *Cities in Perspective I. Economy, Planning and the Environment*, Assen/Nl: van Gorcum, pp. 83–93.

Schamp, E.W. (2000), 'Die Stabilität regionaler Unternehmenscluster – Ergebnisse wirtschaftsgeographischer Forschung' [The stability of regional firm clusters – results of economic geographic research], in H.H. Hirsch-Kreinsen and A. Schulte (eds), *Standortbindungen. Unternehmen zwischen Globalisierung und Regionalisierung* [*Locational Ties. Firms Between Globalisation and Regionalisation*], Berlin: Sigma, pp. 61–101.

Schamp, E.W. (2002), 'Evolution und Institution als Grundlagen einer dynamischen Wirtschaftsgeographie: Die Bedeutung von externen Skalenerträgen für geographische Konzentration' [Evolution and institution as a basis for a dynamic economic geography: the meaning of increasing returns for geographical concentration], *Geographische Zeitschrift*, **90**, 40–51.

Schamp, E.W. (2005), 'Decline of the district, renewal of firms: an evolutionary approach to footwear production in the Pirmasens area, Germany', *Environment and Planning A*, **37**, 617–34.

Schmitz, H. (1995), 'Collective efficiency: growth path for small-scale industry', *Journal of Development Studies*, **31**, 529–66.

Schulz, C., H.P. Doerrenbaecher and C. Liefooghe (2006), 'Far away, so close? Regional clustering of mail order firms and related business services in the Lille metropolitan area', in J.W. Harrington and P.W. Daniels (eds), *Knowledge-Based Services, Internationalization and Regional Development*, Aldershot: Ashgate, pp. 289–305.

Scott, A. and A.S. Paul (1990), 'Collective order and economic coordination in industrial agglomerations: the technopoles of Southern California', *Environment and Planning C*, **8**, 179–93.

Segal Quince & Partners (1985), *The Cambridge Phenomenon. The Growth of the High Technology Industry in a University Town*, Cambridge: Segal Quince & Partners

Setterfield, M. (1995), 'Historical time and economic theory', *Review of Political Economy*, **7** (1), 1–27.

Staber, U. (2007), 'Contextualizing research on social capital in regional clusters', *International Journal of Urban and Regional Research*, **31** (3), 505–21.

Storper, M. (1997), *The Regional World*, New York: Guilford Press.

Storper, M. and R. Walker (1989), *The Capitalist Imperative. Territory, Technology, and Industrial Growth*, Oxford: Blackwell.

Von Tunzelmann, N. (2003), 'Historical coevolution of governance and technology in the industrial revolutions', *Structural Change and Economic Dynamics*, **14**, 365–84.

Zook, M. (2002), 'Grounded capital: venture financing and the geography of the Internet industry', *Journal of Economic Geography*, **2**, 151–77.

21 Locked in decline? On the role of regional lock-ins in old industrial areas

Robert Hassink

When the wind of change blows, some build walls, others build windmills. (Chinese proverb)

1. Introduction

Most of the currently debated theoretical concepts in economic geography try to explain the positive sides of geographical clustering of industries, by emphasizing the positive effects they have on networking and innovation (Lorenzen, 2005). This is exemplified by the work on the rise of high-tech regions, industrial districts and regional production clusters in North America and Western Europe and, to some extent, contrary to older theories on unbalanced regional development that addressed both positive and negative aspects of regional evolution (Myrdal, 1957). This chapter takes a critical stance on the modern one-sided view of the correlation between geographical clustering of industries, networking and innovation. It makes clear that the geographical clustering of industries in some constellations negatively affects innovativeness and renewal. This is particularly the case in the under-theorized old industrial areas (Cooke, 1995; Hamm and Wienert, 1989). To some extent they can be regarded as the industrial districts of the past, in which initial strengths based on geography and networks, such as industrial atmosphere, highly specialized infrastructure, close inter-firm relations and strong support by regional institutions, turned into barriers to innovation. In the 1990s Grabher (1993) emphasized the role of path dependence and lock-ins as concepts explaining the lack of renewal in old industrial areas in general and the Ruhr area in Germany in particular. His studies can be linked to recent work done in order to introduce evolutionary economics into economic geography (Boschma and Frenken, 2006; Boschma and Lambooy, 1999a; Essletzbichler, 2002; Essletzbichler and Winther, 1999; Martin and Sunley, 2006; Maskell and Malmberg, 1999; Morgan, 1997; Schamp, 2000; Shapira and Fuchs, 2005; Storper, 1997; see also Boschma and Martin, and Martin and Sunley, Chapters 1 and 3, this volume). Unlike neoclassical theory, this school takes history and geography seriously by recognizing the importance of place-specific elements and processes to explain broader spatial patterns of technology evolution. Since the lock-in concept is one of the few promising modern concepts explaining the negative sides of clustering, it forms the theoretical core of this chapter. However, the lock-in concept has some weaknesses I aim to tackle here. First, the inductive, bottom-up concept is too weak to be a sound theoretical concept in economic geography yet, as it is mainly based on observations in a few regions (the Ruhr area mainly) and one kind of traditional industrial complex (steel and coal-mining). Second, little systematic cross-regional and cross-sectoral comparative research has been done on the specific role of lock-ins in hindering restructuring processes in old industrial areas. There are a few recently published exceptions, such as the papers by Chapman (2005), Schamp (2005), Hudson (2005) and Van Grunsven and Smakman (2005), published in a special issue of *Environment and Planning A*. This chapter is written in line with this new empirical research strand (see also Birch et al., 2010; Hodson, 2008; Tödtling and Trippl, 2004).

The first aim of this chapter is therefore to comparatively analyse the impact of lock-ins on restructuring processes in two differently structured regions (textile and shipbuilding industry), regions that are furthermore located in differently politically structured countries with a dissimilar economic development level, namely Germany and South Korea. The second aim is to work out factors explaining why it is that we find relatively strong lock-ins in some old industrial areas and relatively weak ones in other old industrial areas. The deliberate choice of starkly contrasting characteristics of research objects is related to the expected economic-structural and political-institutional *impact factors* contributing to the strength of regional lock-ins. Concerning the economic-structural impact factors the *expectation* is that the shipbuilding industry clearly has stronger tendencies towards regional lock-ins than the textile industry, given its stronger spatial concentration and mono-structure, high entry and exit barriers because of its capital-intensive characteristics and its oligopolistic market structure. Concerning the political-institutional impact factors, it is expected that the German associative model is expected to lead to a stronger involvement of local and regional actors involved in lock-ins than in the Korean developmental state model. Creating clear value added to existing individual case-studies (Cho and Hassink, 2009; Eich-Born and Hassink, 2005; Hassink, 2007b; Hassink and Shin, 2005b), this chapter goes two steps further by, for the first time, systematically comparing the four case studies and by, also for the first time, working out and testing the explanatory value of impact factors.

In the next theoretical section on regional lock-ins, these impact factors are presented. After that section, the empirical case studies are put in an economic-structural and political-institutional context in section 3. The case studies are described in section 4, whereas section 5 presents a comparative analysis and draws the conclusions of this chapter.

2. Regional lock-ins in old industrial areas: a theoretical framework

Path dependence and lock-in are important notions of evolutionary economics that have been used by economic geographers to explain the negative sides of economic clusters, particularly the decline in old industrial areas (Boschma and Lambooy, 1999b; Hassink and Shin, 2005a; for an excellent overview of path dependence and regional lock-in and the links to evolutionary economics, see Martin and Sunley, 2006). According to Saxenian (1994, p. 161), 'spatial clustering alone does not create mutually beneficial interdependencies. An industrial system may be geographically agglomerated and yet have limited capacity for adaption. This is overwhelmingly a function of organizational structure, not of technology or firm size'. Geographically concentrated clusters can become insular, inward-looking systems, as many old industrial areas, both resource-based mono-structural areas, dominated by for instance steel, coal-mining and shipbuilding industry, and areas specialized in consumer goods (textile for instance) (Schamp, 2000), have shown us (Hamm and Wienert, 1989; Hudson, 1994). 'The initial strengths of the industrial districts of the past – their industrial atmosphere, highly developed and specialized infrastructure, the close interfirm linkages, and strong political support by regional institutions – turned into stubborn obstacles to innovation' (the 'rigid specialization' trap) (Grabher, 1993, p. 256).

Grabher (1993) has defined these obstacles as three kinds of lock-in. First, a functional lock-in refers to hierarchical, close inter-firm relationships, particularly between large enterprises and small and medium-sized suppliers, which may eliminate the need for suppliers to develop so-called boundary spanning functions, such as research and development and marketing. The lack of these functions hinders suppliers in switching to new markets in times of a structural crisis. Second, a cognitive lock-in is regarded as a common world-view or mindset that might confuse secular trends with cyclical downturns. Third, and closely related to cognitive lock-ins is the notion of political lock-ins that might come up in a production cluster (Grabher, 1993; Hamm and Wienert, 1989; McGillivray, 2004). Political lock-ins are thick institutional tissues aiming at preserving existing traditional industrial structures and therefore unnecessarily slowing down industrial restructuring and indirectly hampering the development of indigenous potential and creativity. Institutional tissues consist both of networks of organizations, such as political administrations at all spatial levels, trade unions, large enterprises and business support agencies, and things that pattern behaviour, such as norms, rules and written and unwritten laws. With regard to the latter part, there seems to be, therefore, a strong relationship between cognitive lock-ins and political lock-ins. Such a particular and thick institutional tissue can, together with the firms and workers, form a so-called self-sustaining coalition (Grabher, 1993; Hudson, 1994). In such a situation, large companies might not want to give up sites for the attraction of inward investment, as they are afraid to lose qualified employees to competitors. Local authorities might not see the point of attracting inward investment or promoting restructuring in another way, as large tax incomes are paid by traditional industries. In some regional production clusters, the spirit of the Schumpeterian entrepreneur might dwindle because of increasing industrial concentration and the domination of large companies. The self-sustaining coalition also lobbies for sectoral interventions often at a national or supra-national level, which hamper the restructuring process more than they support it, as they remove the incentives to take initiatives for entrepreneurs and thus paralyse competition and tranquillize large industries (Hamm and Wienert, 1989). Morgan and Nauwelaers (1999) emphasize that in these kinds of network status is privileged over knowledge, power over learning and the past over the present.

Taken together these three forms of lock-in, functional, cognitive and political, can be considered as *regional lock-ins* (see also Martin and Sunley, 2006 and Boschma, 2005), which forms the core concept in this chapter. A regional lock-in refers to a set of interrelated lock-ins that manifest themselves at the regional level, but are influenced and affected by both intra-regional and extra-regional factors. In a way regional lock-ins explain why we can find in some mature industry clusters *adjustment*, 'which refers to an extension of established trends, resulting in stagnation or gradual decline' or a lack of *renewal*, which would involve 'a significant change of the existing trajectory of development, enabling the cluster to sustain its prosperity' (Chapman et al., 2004, p. 383). In the case of adjustment, firms tend to focus on cost reduction and copying, whereas in the case of renewal, the focus will be on innovation and diversification. If institutional resistance to restructuring is strong in old industrial areas suffering from de-industrialization (strong cognitive and political lock-ins), there is a strong tendency for conserving existing structures or for modernizing existing production facilities (adjustment). If institutional resistance to restructuring is weak, there might be more room for setting up new industries, partly emerging out of the existing industries (renewal), although this is no deterministic relationship, as also in a situation of weak resistance there might be no evolution of new industries. Related to the issue of new industries are the recent discussions on the role of related/unrelated variety and path creation in restructuring regional economies (Boschma and Wenting, 2007; Frenken et al., 2007; Martin and Sunley, 2006). On the one hand, variety is seen as a source of regional knowledge spillovers, measured by *related* variety within sectors. On the other hand, in the case of *unrelated* variety, variety is seen as a portfolio protecting a region from external shocks. According to Martin and Sunley (2006, p. 421) 'there is a trade-off between specialization and a short-lived burst of fast regional growth on the one hand, and diversity and continual regional adaptability on the other'. In most regional economies, however, the situation is rather complex, as:

> various networks and structures of interrelatedness can emerge between different sectors and activities within a region, thus suggesting the possibility of what we might call 'path-interdependence', that is situations where the path-dependent trajectories of particular local industries are to some degree mutually reinforcing. The extent and significance of this interlinking path effect is a key issue for further research. (Martin and Sunley, 2006, p. 421)

Although it is acknowledged that related and unrelated variety plays a role in the overall restructuring of a regional economy, the main focus in this chapter is on the role of lock-ins in the hindering of restructuring processes.

Several authors suggest that mono-structural regional economies with a high degree of specialization, in particular, are most prone to regional lock-ins (Grabher, 1993; Hamm and Wienert, 1989; Martin and Sunley, 2006; Schamp, 2000, p. 136). Going one step further, it is also suggested that regional lock-ins are relatively strong in spatial concentrations of capital-intensive industries, such as the steel, coal-mining and shipbuilding industries, which are spatially more concentrated than labour-intensive traditional industries, such as textiles, and which are often characterized by high degrees of state involvement at national and supra-national level leading to stronger protests and resistance in case of politically influenced plant closures (Hamm and Wienert, 1989; Hudson and Sadler, 2004, p. 291; Schamp, 2000). It is also emphasized that regional lock-ins are embedded in varying national and supra-national institutional contexts (Hudson and Sadler, 2004; Martin and Sunley, 2006; Schamp, 2000, p. 145).

The line, however, between adjustment and renewal cases can be very thin (Boschma and Lambooy, 1999b; Grabher, 1993; Hamm and Wienert, 1989; Martin and Sunley, 2006; Maskell and Malmberg, 1999; Tödtling and Trippl, 2004). This is illustrated by Essletzbichler and Winther (1999) when they speak about positive and negative lock-ins, by Fürst and Schubert (1998) when they distinguish between productive and non-productive political networks and by Callon (1998) when he refers to enabling and constraining networks. As milieus tend to change more slowly than industries, a sclerotic milieu can remain in a region even after the industrial structure to which it belonged already has disappeared. The transition from positive to negative lock-ins and the thin line between 'good' and 'bad' industrial agglomerations (Hassink, 1997; Saxenian, 1994) show the importance of studying and understanding this under-researched phenomenon in economic geography. In this chapter, the constraining and negative character of lock-ins is emphasized, as it analyses the explanatory role lock-ins can play in delaying necessary renewal processes in old industrial areas. A sound understanding of negative lock-ins is for several reasons important for geographers. They might explain the structural economic problems some old industrial areas face, as well as the related persistence of regional economic inequalities in some industrialized countries. Moreover, a sound understanding of the emergence of negative lock-ins might enable geographers to draw policy lessons on how to avoid their emergence.

Thus, the evolutionary economics school and the related lock-in concept seem to be useful concepts to understand the negative consequences of path-dependent development and the importance of regions' capabilities to adjust their institutional endowments ('un-learning') (Maskell and Malmberg, 1999; Schamp, 2000). Grabher's lock-in concept has been often cited (see for instance in Cooke and Morgan, 1998, p. 111; Schamp, 2000, p. 139), showing its importance as an explanatory concept for the decline of industrial areas, but empirical research testing the concept has been rare.

Moreover too little is known to answer the key question, namely 'why it is that some regional economies become locked into development paths that lose dynamism, whilst other regional economies seem able to avoid this danger?' (Martin and Sunley, 2006, 395). This chapter, therefore, not only aims at analysing whether we find regional lock-ins in old industrial areas, but also tackles this key question by looking at factors having an impact on regional lock-ins. Derived from the above discussions on path dependence, lock-ins, institutions and industrial characteristics, the following impact factors are expected to have the strongest effect on regional lock-ins. One could of course debate adding more factors, such as the age of an industry, but these are regarded as the most important ones. Impact factors are not the same as environment or context; they also include the characteristics of the population, as can be seen under the economic-structural impact factors. The strength of regional lock-ins, in turn, affects adjustment or renewal in old industrial areas (Figure 21.1).

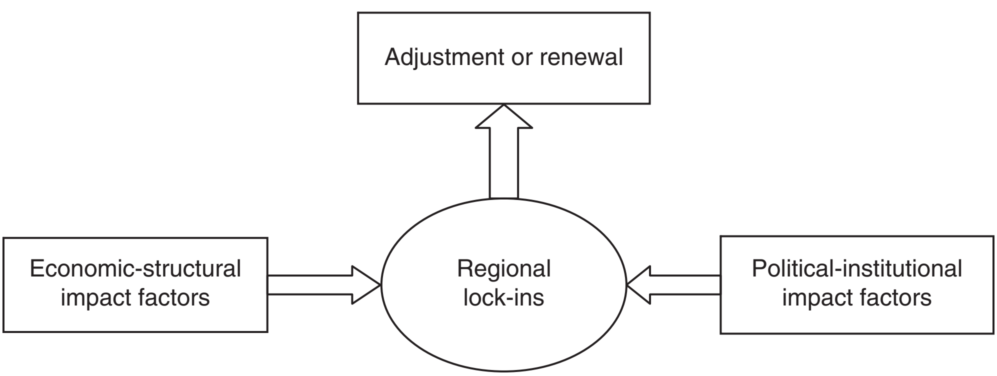


Figure 21.1 Relations between impact factors, regional lock-ins and the restructuring in old industrial areas

Economic-structural impact factors

1. A marked industrial mono-structure: the leading industry having an employment share of at least 30 per cent of the total manufacturing employment in the region as a rough indicator for a mono-structure.
2. A specific leading industry: capital-intensive, high entry and exit barriers, above average company size, oligopolistic market structure, and influential trade unions.

Political-institutional impact factors

3. An institutional tissue at the regional level, consisting of local, regional policy-makers, captains of industry, regional trade unionists, representatives of industry associations, that is strongly focused on the leading industry and hence weakly on external relations.
4. A national-political system, which enables regional actors to influence political questions concerning industrial policy.
5. Supra-national institutions that strongly affect the conditions of industrial policy relevant to the leading industry.

The following empirical section analyses whether regional lock-ins hinder the renewal of the textile and shipbuilding industry regions in Germany and South Korea. It also helps to fulfil the second main aim of this chapter, namely to detect the factors affecting the strength or weakness of regional lock-ins.

3. Putting the case studies in context

In order to be able to give a fair judgement on the impact of the above described factors on regional lock-ins, the case studies that are presented in the next section need to be contextualized both concerning the economic structure (shipbuilding and textile industries) and the political-institutional context (Germany and South Korea).

Concerning the *economic-structural context*, the shipbuilding industry is both spatially extremely concentrated, and at the same time a global industry and an industry with much state intervention (Cho and Porter, 1986; Stopford, 1997). During the history of the industry, there have been dramatic changes in global leadership (Cho and Porter, 1986; Stopford, 1997; Todd, 1991). Shortly after the Second World War, Germany and some other European countries took over leadership from Great Britain. In the 1960s Japan became the world's leading shipbuilding nation. Since 1973 South Korea has been building up and expanding its shipbuilding industry and since the end of the 1990s Japan and South Korea have shared world leadership in shipbuilding (Stopford, 1997). The industry is characterized by very high exit barriers and high sunk costs. Furthermore, in some countries the industry is strongly supported by the state because many shipyards are located in coastal areas with few alternative employment opportunities (Stopford, 1997, p. 468). 'These [exit] barriers lead unsuccessful shipbuilders to continue operating, often with government subsidies, and to persistently wage price wars at the expense of industry profitability' (Cho and Porter, 1986, pp. 543–4). Despite high entry barriers, strong government support has facilitated several countries, such as Japan and South Korea, to enter the market, whereas governments in Western Europe heavily intervened in order to protect the industry against competition from low-wage countries (Stråth, 1986). Related to these government policies, the European and South Korean shipbuilding industries have recently been involved in a political battle (Eich-Born and Hassink, 2005).

The textile industry, in contrast, is much less spatially concentrated, less capital-intensive, has lower entry and exit barriers, is more characterized by small and medium-sized enterprises, has less influential labour unions and has in many countries, notably in Germany and South Korea, a much weaker lobby for state support (Dicken, 2003; Toyne et al., 1984).

Despite the competitive pressure and economic globalization, the national *political-institutional context*s of development remain remarkably distinct (Whitley, 2000) and therefore still provide different institutional contexts for lock-ins. Three different socio-economic models of development can be identified (see also Cooke and Morgan, 1998; Whitley, 2000): entrepreneurial, associative and developmental. Germany can be considered as an associative and South Korea as a developmental model.

The *associative model* of development is often regarded as a model of 'public–private partnership', where states are negotiating with, and often also delegating power to, industrial associations. Industrial associations have emerged as important vehicles for branch organization, concerned with standardization or other topics of common interest, also influencing market regulations and the forms of competition. The state enables forms of self-organization that foster both economic development and political stability (Cooke and Morgan, 1998). This form of governance involves the devolution and delegation of power from the state to intermediate organizations, thus creating institutions that may enhance and enable learning and innovation. Thus, the state is collaborating with and working through institutions that constitute systems of innovation (Cooke and Morgan, 1998, p. 24). The associative model is typically found in Germany, Austria and Switzerland, partly in the Netherlands and in a different, participatory form in Scandinavian countries.

The associative model has a broad-based educational and training system that recognizes the value of both academic education and vocational training (Lam, 2002). This educational system is conducive to a decentralized mode of work organization and provides a good basis for interactive learning. Practical experience has a high social status. Although this educational system allows for the accumulation of tacit knowledge, at the same time it inhibits the creation of active labour markets. Labour markets are much less polarized than in the entrepreneurial model. They are partly local, with relatively little geographical mobility and scope for renewal. The associative model is characterized by strong unions and participation of the work-force in management, which leads to long-term relations to employers.

With regard to the constitutional set-up, the state encourages the establishment of strong intermediary associations and delegates a considerable range of economic and social functions to them (Whitley, 2000). There is strong involvement of social partners in policy-making. Associative modes of governance also play an increasingly important role in regions, transforming the regional level from being passive spaces subjected to corporate allocative decision-making into laboratories where regional institutions enhance trust and cooperation and have contributed to innovation processes (Cooke and Morgan, 1998). That means, on the other hand, that local and regional actors also have more leeway to resist necessary restructuring processes and hence regional lock-ins are potentially stronger.

The *developmental state model* is the East Asian variant type of the broader state-centred socio-economic model of development, which can be found in France (see Cooke and Morgan, 1998). In order to close the industrial and technological gap with the West, Japan and South Korea followed the route of the developmental or plan-rational state, characterized by a strong, authoritarian, central government that deliberately and strategically supported large enterprises and competitiveness in certain selected targeted industries. Economic policy successfully followed the sequence of import, import-substitution and export orientation.

Both countries lack natural resources and are consequently strongly committed to education and the development of human resources. The educational system is a combination of broad-based training system and elite academic system. There is recognition of the value of both academic education and vocational training. A high value is put on practical skills of engineers, partly for historical reasons: industrial development was based on imported technology and hence engineers played a crucial role in international technology transfer. The labour market is characterized by long-term stable employment in large enterprises.

The corporate governance of the developmental model is characterized by a high level of state-controlled ownership coordination (Whitley, 2000). The state strongly controlled the capital market; banks were mostly state-owned and provided biased financial support for targeted large enterprises. South Korea's large conglomerates (chaebol) were strongly controlled by the state and tended to be characterized by strong vertical integration processes and a centralized, hierarchical and kinship-based organization.

The constitutional set-up is characterized by a strong central government and its ministries and agencies. The combination of a strong central state and chaebol with their headquarters in Seoul gives local and regional actors little leeway to affect restructuring processes and hence regional lock-ins might be rather weak.

Recently, however, a clear development can be observed from the developmental state model in crisis towards the entrepreneurial socio-economic model of development, a development that particularly started after the economic crisis of 1998 (Moon and Rhyu, 2000). This shows that the presented models are dynamic and change through time and that their influence on regional lock-ins can be ambiguous and should hence be carefully monitored.

The above-described economic-structural context and political-institutional context lead to the following expectations: compared to the textile industry, the shipbuilding industry clearly is expected to have stronger tendencies towards regional lock-ins, given its stronger spatial concentration and mono-structure, high entry and exit barriers because of its capital-intensive characteristics and its oligopolistic market structure. The German associative model is expected to lead to a stronger involvement of local and regional actors involved in lock-ins than in the Korean developmental state model.

4. Locked in decline? Empirical case-studies

In the following, the four case-studies are presented, including the textile industry region Westmünsterland and the shipbuilding region Mecklenburg-Vorpommern, both in Germany, and the textile industry region Daegu and the shipbuilding region Gyeongnam,

Figure 21.2 Location of Mecklenburg-Vorpommern and Westmünsterland (Steinfurt and Borken) in Germany

both in South Korea (see Figures 21.2 and 21.3 for locations). Data were collected by studying literature and policy documents and by conducting in total 83 interviews over a study period of three years with textile and shipbuilding company managers and their suppliers, local, regional, national and supra-national policy-makers, academic experts, consultants, trade unionists and officials of industry associations.

The shipbuilding region of Mecklenburg-Vorpommern, Germany

Shipbuilding dominates the regional production structure of Mecklenburg-Vorpommern, Germany (Eich-Born, 2005; Eich-Born and Hassink, 2005), one of the new Länder in reunited Germany situated in the North East (Figure 21.2). Because of the transformation from central planning to market economy, de-industrialization (employment in shipbuilding dropped from 55,000 in 1989 to around 5000 in 2003) led to dramatically high unemployment rates of around 20 per cent. In order to save the industry from total collapse, a political consensus was built between various interest groups on different geographical levels: yard managers, workers' councils, regional trade unions, mayors of yard cities, regional policy-makers as well as the representatives of the German

Shipbuilding and Ocean Industries Association and the Coordinator for the Maritime Economy in the Federal Ministry of Economics and Technology. Dissolution of the state-owned enterprise and privatization were the strategies applied by the German national and regional government. Nowadays the yards are in the hands of mainly Scandinavian shipbuilding concerns. The modernization of the production capacities was mainly financed by subsidies provided by the German government, the state of Mecklenburg-Vorpommern and the European Union. Over a period of five years, the German government invested more than DM 6 billion (€3 billion) in the construction of new docks, which means state support of about DM 1 million (€500,000) per job (Röller and von Hirschhausen, 1996, p. 17). For each DM of state aid only about 0.09 DM of private investment was attracted. In order to avoid a strong increase in over-capacity, the federal government and the European Commission agreed that the shipyards in Mecklenburg-Vorpommern were not allowed to build more ships than a certain annual capacity limit until 2005.

Thus, regional industrial policy in Mecklenburg-Vorpommern is very much focused on preserving the existing shipbuilding complex, rather than focused on developing new products and industries. That is not only shown by the large amount of subsidies made available for the shipbuilding industry, it is also shown by the successful lobby of the consensus group of actors to release the EU capacity limitation and to sue the South Korean government at the WTO for supposedly illegally supporting Korean yards. It would be unfair, however, to make the impression that the regional industrial policy is only active in lobbying activities in order to preserve existing structures. It also does a lot to support innovative small companies and innovation projects. However, these projects, such as the Maritime Alliance in the framework of the federal support programme InnoRegio, also mainly support the innovativeness of the existing cluster. All in all, both the industrial policy and the regional innovation policy strongly focus on adjustment rather than renewal. We can therefore clearly observe a 'noisy' restructuring process because of relatively strong regional lock-in, which can partly be explained by the clear mono-structure and the lack of alternative industrial activities and by the specific characteristics of the leading industry (capital-intensive large enterprises, so only a few main actors in the region). But at the same time also political-institutional factors also play their role, such as a thick and powerful institutional tissue at the region level in combination with a federal government determined to support the industry.

The shipbuilding region of Gyeongnam, South Korea

Korea has become a world market leader in shipbuilding within just a time period of about 30 years (Hassink and Shin, 2005b; Woo, 2003). The industry has been virtually built up from scratch, as Korea's world market share in the early 1970s was about 2 per cent, compared to the current 38 per cent. Shipbuilding started in the early 1970s because of the interventionist developmental state led by Park Chung-hee (Amsden, 1989). The government's extensive control over the financial sector enabled channelling of investment funds to the industry, which has been heavily concentrated organizationally in a few extremely large shipyards (the world's three largest shipyards are located in Korea) and geographically in the province of Gyeongnam (Figure 21.3). Apart from Halla, which is located in Mokpo (Jeonnam), Korea's seven main shipbuilding companies have their manufacturing facilities in the region (including Ulsan and Busan), the largest being

Figure 21.3 Location of Gyeongnam and Daegu in South Korea

in Ulsan (Hyundai Heavy Industries) and Geoje (Daewoo and Samsung); other large yards can be found in Ulsan (Hyundai Mipo), Busan (Hanjin) and Jinhae (STX).

Both academics and the popular press in Europe are very much focused on the role of the central government in subsidizing Korea's shipbuilding industry, as an explanation for its success. The industry, however, has developed from a couple of cathedrals in the desert, highly dependent on import for its main supply, towards an innovative cluster with much interaction between the large yards and their suppliers, as well as with universities and research institutes (Hassink and Shin, 2005b). The production complex developed from simple co-location enforced by central government and favoured by the natural physical conditions in the area, as well as favouritism by political leaders of central government, to a full fledged cluster, with strong competition, collaboration and policy support at central level. Until recently, local and regional governments played a minimal role in supporting the cluster.

Presently, the shipbuilding cluster, which mastered the financial crisis in 1998 remarkably well, is enabling regional growth. However, the growth-enabling lock-in in Gyeongnam might turn into a constraining one in the future, if external conditions negatively affect the cluster and at the same time lock-ins hinder the ability of the cluster to react to the changing conditions. Among these changing conditions are the conflicts with the European Union on unfair competition, the expected cyclical downturn in the industry because of over-capacity combined with the increasing competitiveness of low-cost shipbuilder China, which might increasingly make shipbuilding in South Korea a vulnerable industry and which might turn the cluster from a mature into a declining one. Potential lock-ins might occur because of the common worldview shared by the main actors in the industry that China is not yet a threat (cognitive lock-in), and because of the high dependencies between suppliers and large yards in the close-knit cluster that is still completely in Korean hands (functional lock-in). If, therefore, competition from China strongly increases in a future scenario, a constraining lock-in may occur. This constraining lock-in might, however, be much more the product of national interests than interests of a regional coalition of enterprises and local and regional policy-makers.

The textile region of Westmünsterland, Germany

Thinly populated Westmünsterland, consisting of the counties Steinfurt and Borken, is located in North Rhine-Westphalia in the western part of Germany at the Dutch–German border (see Figure 21.2). Within Germany it is one of the main textile clusters. The region's textile industry started with linen-weaving in cottages and farmhouses and was initially based on flax and hemp (Ditt, 2000; Hassink, 2007b; Hauff, 1995). Factory production started from the 1840s onwards, when spinning machinery was adopted. Most firms that were established were small and family-owned. At the beginning of the 1960s, in its heyday, the textile industry employed about 75 per cent of all employees in the region's manufacturing industry. In the early development phase, local textile entrepreneurs had a strong impact on local politics. The exact location of railroad tracks, self-financed railroad construction and the establishment of weaving schools are examples of the entrepreneurs' impact on shaping the local production environment (Hauff, 1995, p. 135). During its heyday the textile cluster in Westmünsterland had the characteristics of an industrial district, that is intensive inter-firm networks within the region, but because of the internationalization of sales, procurement and recently also production, the intensity of inter-firm networks within the region has been rapidly decreasing.

Because of a strong increase in competition from emerging economies in East Asia combined with automation and rationalization of the production process, textile employment decreased by 25 per cent during the 1960s. In the subsequent 1970s and early 1980s Westmünsterland's economy hence suffered from a crisis. However, it now has scores that are average or even above-average on socio-economic indicators such as employment growth, unemployment and gross domestic product per capita. The restructuring process has led to a relatively heterogeneous regional production structure. Instead of dominating the regional production structure, the textile industry is now only one of a few large industries. The remaining textile companies have been successful in reorientating themselves towards technical textiles, the high value-added segment of the textile market. Inter-firm cooperation within the region has been decreasing to low levels, as a result of the restructuring process and the focus of the remaining textile companies on new, diverse product markets, which are located outside the region. Overall, the regional economy has been relatively successfully restructured (Reckfort and Ridder, 1996), since new manufacturing and service employment could be created to compensate for the job losses in the textile industry. Moreover, most of the remaining textile companies managed to achieve renewal by diversifying into the field of technical textiles, with applications in, for instance, medical technology, the car supplier industry, construction and so on, instead of adjustment through copying and cost reduction.

It is partly because of relatively weak functional, cognitive and political lock-ins that this renewal could be achieved. They contrast with the relatively strong lock-ins observed in the neighbouring heavy industrial complex of the Ruhr area. Weak functional lock-ins can be seen through the early diversification of the few textile industry suppliers in the region into new markets. Weak cognitive and political lock-ins are testified by the early mental and policy reorientation of the main regional actors towards supporting new economic activities. This region has been going through 'quiet' restructuring processes, which, in turn, led to renewal and a diversification of the production structure. Political lock-ins and resistance to restructuring were weak in this region and closely related to that, there tended to be little media and academic attention for its economic problems (Hauff, 1991, 1995). Hauff (1991, p. 207), for instance, observed at the beginning of the 1980s, after the collapse of Van Delden, the largest textile firm in the region:

> In the wake of the Van Delden collapse, Gronau's unemployment rate was, in fact, higher than that of any other urban community in North Rhine-Westphalia. It is, indeed, surprising that, despite the tenseness of the job situation in and around Gronau, there were neither protest movements, nor occupations of factories like those which had accompanied pit and steel closures in the more militant Ruhr.

This 'quiet' restructuring process and related weak regional lock-ins can be explained by factors that are partly industrial and partly contextual in nature. With regard to the first factor, relatively weak lock-ins can be explained by the specific characteristics of the textile industry, an industry with, in contrast to heavy industrial complexes, low entry and exit barriers, relatively many small and medium-sized enterprises, strong competition and little influence of trade unions (Dicken, 2003). Furthermore, a number of specific contextual factors can be observed in Westmünsterland that contributed to weak lock-ins. First, the location of the region adjacent to the heavy industry complex of the Ruhr area, within the political-administrative boundaries of one state (North Rhine-Westphalia), led to little media attention and weakened the hopes of the main regional actors to be able to successfully lobby the state government. Second, local cultural and organizational factors, such as the reserved, down-to-earth and settled mentality of its people and the patriarchal organization of family firms (Hassink, 2007b), contributed to weak lock-ins. Because of the weak lock-ins, the regional economy reoriented itself at an early stage, which in turn led to a successful renewal.

The textile region of Daegu, South Korea

The textile industry cluster of Daegu, the third largest city in South Korea (see Figure 21.3), and the surrounding cities of Gumi and Gyeongnam, started to grow in the 1960s (Cho and Hassink, 2009; Lee et al. 2000; Park, 1997). It is characterized by specialization in the production and weaving of chemical fibres and has been strongly focusing on export. Textile business constitutes the largest segment of manufacturing in Daegu: 31.3 per cent of total establishment, 34.7 per cent of total employment; which means a location quotient of 4.1, 34.6 per cent of total production, 54.2 per cent of total export and 30.9 per cent of total value added in 2002 (Cho and Hassink, 2009). The high rate of automation in the 1980s brought about problems of overcapacity and overproduction, which in turn led to financial difficulties in the textile business. Moreover, textile companies were faced with increasing competition from low-cost neighbouring countries, China, in particular, and a shift of Korean producers to China in the 1990s. Forty years of path-dependent evolution led to specialization in the narrow low value added and low-tech middle stream of the textile value chain, whereas high value added and high-tech downstream activities were nearly totally absent. The upshot of all this is that the employment in Daegu's textile industry decreased from 91,000 in 1981 to 82,000 in 1986 and 47,000 in 2000 (Cho and Hassink, 2009).

In the context of this situation, Cho and Hassink (2009) observed both functional and cognitive lock-ins. As for the functional lock-in, there are a large number of small-scale vertically as well as horizontally interrelated producers around a handful of large supplier firms of 'raw fibres' and a number of large textile companies with market outlets. Functionally, the networks are geared to producing, weaving and dyeing chemical fibres, typically constituting a middle-stream segment in the textile industry. This means that Daegu's textile production system is based on hierarchical cost-cutting and subcontracting networks, of which the driving force is the sustained demand on foreign markets. Much of the high valued added part of the textile value chain, such as apparel, design, fashion and trade, however, has been displaced or integrated into extra-local networks in Seoul, with Daegu's textile industry falling into a trap of low value added production. Concerning the cognitive lock-in, local clientelism gives rise to passive and captive behaviour of local entrepreneurs to exogenously imposed conditions for textile production. Enjoying export demand and policy protection, small local textile producers stick to their vested interests to avoid risk-taking ventures. This commonly shared vision has led to local mercantilistic entrepreneurship among second-generation local entrepreneurs. They tend to seek rents from subcontracting, land speculation and factory leasing rather than to generate profits from high value added production. They are not interested in improving their learning capability and act merely as a self-sustaining coalition that resists a progressive reshuffling of the existing industrial structure. Because of this mercantilistic entrepreneurship, the existing industrial networks of middle-stream textiles are clearly cognitively locked in.

In response to this situation, the central government launched an ambitious project called the Milano Project (1998–2003) aiming at restructuring the present middle-stream textile of Daegu into a high value added down-stream textile that comprises apparel, design and fashion as a competitive edge. Milano is a symbolic target for the high-end restructuring of textile in Daegu. In April 1998, President Kim Dae-Jung, who came to power with regionalist full support from his home province in the south west, visited Daegu to lessen the south east's regionalist antagonism against him and officially promised (*kongyak*) full policy support for revitalization of Daegu's decaying textile industry, which materialized into the Milano Project. It consists of 19 projects in four sectors, which require a total of €650 million in a period of five years. As of 30 April 2003, the overall rate of project implementation was 75 per cent. The main promoters of the project, the central government and the City of Daegu, aim at promoting both new activities (fashion and design) and projects with new actors (research institutes, universities, design schools, banks etc.), whereas the actors with a vested interest, local textile producers and their lobby organizations, oppose these plans. The latter argued that Daegu's textile cluster should maintain its competitive edge in the branch of weaving and dyeing, whose technology, know-how and market accessibility were believed to be at the top of the world. Therefore resistance to and conflict around the restructuring are widely witnessed in the process of project implementation and the project is, therefore, not regarded to be a success (Cho and Hassink, 2009; Sohn, 1999). In a way, therefore, centrally devised renewal efforts were blocked by locally emerged lock-ins.

5. Comparative analysis and conclusion

Are old industrial areas locked in decline? Not necessarily so is the first conclusion that can be drawn after analysing restructuring processes in four regional economies. The chapter has shown that some regional economies, such as Westmünsterland, successfully renewed and diversified their economies, partly because of the lack of regional lock-ins and hence the 'quiet restructuring' process. Mecklenburg-Vorpommern and Daegu, on the other hand, are locked in adjustment, partly because of strong lock-ins blocking necessary renewal (Table 21.1). Analysing the impact of lock-ins on restructuring processes was the first aim of this chapter.

The second aim formulated in the introduction of this chapter was to work out *impact factors* to explain the differences found: why is it that we find relatively strong regional lock-ins in some old industrial areas and relatively weak ones in other old industrial areas? Derived from the literature, two types of impact factors were listed, that is economic-structural impact factors and political-institutional impact factors. It was expected that in comparison to the textile industry, the shipbuilding industry clearly has stronger tendencies towards regional lock-ins, given its stronger spatial concentration and mono-structure, high entry and exit barriers because of its capital-intensive characteristics and its oligopolistic market structure. At the same time, it was expected that the German associative model leads to a stronger involvement of local and regional actors involved in lock-ins than in the Korean developmental state model.

The empirical results show that each individual case study can only be explained by a unique set of impact factors (Table 21.1). Mono-structure does not necessarily lead to regional lock-ins, as the case of Westmünsterland shows us. Also the industrial structure does not in all cases explain the differences in lock-ins, as can be seen in South Korea, where a lock-in prone industry such as shipbuilding does not show regionally induced lock-in tendencies, whereas the regionally induced lock-in unexpectedly emerged in the textile region in Daegu. Regionalism and national political factors that led to clientelism play a role here. To understand, therefore, why the intensity of lock-ins differs between

Table 21.1 Comparative analysis of case studies

	Adjustment/ renewal	Regional lock-ins	Dominant impact factors
Mecklenburg-Vorpommern	Adjustment	Strong at several spatial levels	Both economic-structural and political-institutional factors
Gyeongnam	Slow renewal	Fair and enabling, mainly at national level	Economic-structural factors dominate
Westmünsterland	Renewal	Weak	Both economic-structural and specific regional political-institutional factors
Daegu	Adjustment	Strong at local level	Political-institutional impact factors dominate

regional settings, contingent path dependence and context specific factors need to be taken into account for each individual case. These results fit well in the current institutional, relational and evolutionary paradigms of economic geography, which all emphasize the role of path dependence and contingency (Martin and Sunley, 2006). They also confirm the main argument expressed by Martin and Sunley (2006, p. 414), namely that 'we need to understand regional "lock-in" as a multiscaled process, and one which also has a high degree of place-dependence, rather than as a universal principle that applies everywhere and anywhere and that is inexorable in its emergence and consequences'. In order to forecast where lock-ins could block regional renewal in the future, it is thus of utmost importance to go beyond the narrow spatial focus on the local and regional, from which many studies of old industrial areas have been suffering. The empirical part of this chapter has shown that it is of key importance when analysing regional lock-ins in old industrial areas to take the institutional context at all spatial levels, that is local, regional, national, and supra-national into account.

This chapter has attempted to contribute to the key question of why it is that 'some regional economies become locked into development paths that lose dynamism, whilst other regional economies seem able to avoid this danger' (Martin and Sunley, 2006, p. 395). More research is necessary to answer this question and a future research agenda should, in my view, focus on several issues that could not be enough dealt with here. First, not enough research has been done yet on the role of social capital and trust in relation to regional lock-ins. They might potentially, for instance, be strong in regions with high stocks of social capital and trust. Second, the issue of sunk costs (costs that are irrevocably committed to a particular use, and therefore are not recoverable in case of exit) should be more explored in relation to regional lock-ins (Melachroinos and Spence, 2001). Third, too little is known about the issues around related and unrelated variety and path creation in relation to the above-mentioned key question (see also Martin and Sunley, 2006). Fourth, further research is necessary regarding what triggers the transition from positive to negative lock-ins in regions. Last but not least, more work should be done on regional policy concepts to avoid the emergence of lock-ins, such as the learning region and learning cluster (Hassink, 2007a; OECD, 2001). Particularly the paradigm of evolutionary economic geography offers a wide variety of theoretical notes, such as variety, selection environment, path dependence and lock-ins that are useful in dealing with these issues and hence in coming closer to a satisfying answer to the above-mentioned key question in economic geography.

Acknowledgements

This chapter presents the main results of a three-year comparative research project on the impact of lock-ins on old industrial areas in Germany and South Korea, which was sponsored by the Deutsche Forschungsgemeinschaft (German Research Council, DFG) and the Korea Science and Engineering Foundation (KOSEF) (see Grabher and Hassink, 2004). An earlier version of this chapter was presented at the Fourth European Meeting on Applied Evolutionary Economics (EMAEE) Utrecht, The Netherlands, in May 2005. I would like to thank the interviewees for spending their time with me, Ron Boschma, Maureen McKelvey, Bjørnar Sæther, Oliver Ibert, Bjørn Terje Asheim and Gernot Grabher and four anonymous reviewers for giving useful comments on (parts of) earlier versions of this paper and Hubert Lehnard, Werner Schöttelndreyer, Heiner

Heseler, Chul Woo Lee, Woobae Lee and Dong-Ho Shin for their assistance during the project. The usual disclaimer applies, however.

References

Amsden, A.H. (1989), *Asia's Next Giant: South Korea and Late Industrialisation*, New York & Oxford: Oxford University Press.

Birch, K., D. MacKinnon and A. Cumbers (2010), 'Old industrial regions in Europe: a comparative assessment of economic performance', *Regional Studies*, **44** (1), 35–53.

Boschma, R.A. (2005), 'Proximity and innovation: a critical assessment', *Regional Studies*, **39**, 61–74.

Boschma, R.A. and K. Frenken (2006), 'Why is economic geography not an evolutionary science? Towards an evolutionary economic geography', *Journal of Economic Geography*, **6**, 273–302.

Boschma, R. and J. Lambooy (1999a), 'Evolutionary economics and economic geography', *Journal of Evolutionary Economics*, **9**, 411–29.

Boschma, R. and J. Lambooy (1999b), 'The prospects of an adjustment policy based on collective learning in old industrial regions', *GeoJournal*, **49**, 391–99.

Boschma, R.A. and R. Wenting (2007), 'The spatial evolution of the British automobile industry: does location matter?', *Industrial and Corporate Change*, **16**, 213–38.

Callon, M. (1998), 'Introduction: the embeddedness of economic markets in economics', in M. Callon (ed.), *The Laws of the Markets*, Oxford: Blackwell, pp. 1–57.

Chapman, K. (2005), 'From "growth centre" to "cluster": restructuring, regional development, and the Teesside chemical industry', *Environment and Planning A*, **37**, 597–615.

Chapman, K., D. MacKinnon and A. Cumbers (2004), 'Adjustment or renewal in regional clusters? A study of diversification amongst SMEs in the Aberdeen oil complex', *Transactions of the Institute of British Geographers*, **29**, 382–96.

Cho, D.S. and Porter, M.E. (1986), Changing global industry leadership: the case of shipbuilding', in M.E. Porter (ed.), *Competition in Global Industries*, Boston, MA: Harvard Business School Press, pp. 539–67.

Cho, M.-R. and R. Hassink (2009), 'The limits to locking-out through restructuring: the textile industry in Daegu, South Korea', *Regional Studies*, **43**, 1183–98.

Cooke, P. (ed.) (1995), *The Rise of the Rustbelt*, London: UCL Press.

Cooke, P. and K. Morgan (1998), *The Associational Economy; Firms, Regions and Innovation*, Oxford: Oxford University Press.

Dicken, P. (2003), *Global Shift: Transforming the World Economy*, London: Paul Chapman.

Ditt, K. (2000), 'Wirtschaftlicher Wandel in Textilregionen während des 19. und 20. Jahrhunderts: Die Industrialisierung Minden-Ravenbergs und des Westmünsterlandes im Vergleich', [Economic change in textile regions during the 19th and 20th century: the industrialisation of Minden-Ravenberg and Westmünsterland], *Westfälische Forschungen*, **50**, 293–331.

Eich-Born, M. (2005), *Transformation der ostdeutschen Schiffbauindustrie. Anpassungsprozesse in einem global-lokalen Institutionengefüge*, [The Transformation of the Eastern German Shipbuilding Industry. Adjustment Processes in a Global–Local Institutional Tissue], Münster: Lit Verlag.

Eich-Born, M. and R. Hassink (2005), 'On the battle between shipbuilding regions in Germany and South Korea', *Environment and Planning A*, **37**, 635–56.

Essletzbichler, J. (2002), 'Evolutionäre Wirtschaftsgeographie: Neues Forschungsparadigma oder Sackgasse?', [Evolutionary economic geography: new paradigm in research or dead end?], *Geographischer Jahresbericht aus Österreich*, **59**, 11–30.

Essletzbichler, J. and L. Winther (1999), Regional technological change and path dependency in the Danish food processing industry', *Geografiska Annaler B*, **81**, 179–95.

Frenken, K., F. Oort van and T. Verburg (2007), 'Related variety, unrelated variety and regional economic growth', *Regional Studies*, **41**, 685–97.

Fürst, D. and H. Schubert (1998), 'Regionale Akteursnetzwerke; Zur Rolle von Netzwerken in regionalen Umstrukturierungsprozessen', [Regional actor networks: on the role of networks in regional processes of restructuring], *Raumforschung und Raumordnung*, **56**, 352–61.

Grabher, G. (1993), 'The weakness of strong ties; the lock-in of regional development in the Ruhr area', in G. Grabher (ed.), *The Embedded Firm: On the Socioeconomics of Industrial Networks*, London & New York: Routledge, pp. 255–77.

Grabher, G. and R. Hassink (2004), 'Strukturwandel in altindustriellen Regionen in Deutschland und Südkorea', [The restructuring of old industrial regions in Germany and South Korea], University of Bonn, Final Report to the German Research Council (Deutsche Forschungsgemeinschaft), available at: http://www.giub.uni-bonn.de/grabher/downloads/finalreport_SARIDS.pdf.

Hamm, R. and H. Wienert (1989), *Strukturelle Anpassung altindustrieller Regionen im internationalen Vergleich*

[An International Comparison of Structural Adjustment of Old Industrial Regions], Essen: Rheinisch-Westfälisches Institut für Wirtschaftsforschung.

Hassink, R. (1997), 'What distinguishes "good" from "bad" industrial agglomerations?', *Erdkunde*, **51**, 2–11.

Hassink, R. (2007a), 'The learning region: a constructive critique', in R. Rutten and F.W.M. Boekema (eds), *The Learning Region: Foundations, State of the Art, Future*, Cheltenham, UK and Northampton, MA, USA: Edward Elgar, pp. 252–71.

Hassink, R. (2007b), 'The strength of weak lock-ins: the renewal of the Westmünsterland textile industry', *Environment and Planning A*, **39**,1147–65.

Hassink, R. and D.-H. Shin (2005a), 'Guest editorial: The restructuring of old industrial areas in Europe an Asia', *Environment and Planning A*, **37**, 571–80.

Hassink, R. and D.-H. Shin (2005b), 'South Korea's shipbuilding industry: from a couple of cathedrals in the desert to an innovative cluster', *Asian Journal of Technology Innovation*, **13**, 133–55.

Hauff, T. (1991), 'De-industrialisation, socioeconomic change and revitalisation in the west Münsterland textile area: the case of Gronau,' in T. Wild and P. Jones (eds), *De-industrialisation and new industrialisation in Britain and Germany*, London: Anglo-German Foundation, pp. 199–213.

Hauff, T. (1995), *Die Textilindustrie zwischen Schrumpfung und Standortsicherung: weltwirtschaftliche Anpassungszwänge, unternehmerische Handlungsstrategien und regionalökonomische Restrukturierungsprozesse in der Textilindustrie des Westmünsterlandes* [The Textile Industry between Contraction and Location Protection: Under Adjustment Duress due to Worldwide Economic Change, Company Management Strategies and Regional Economic Restructuring Processes in the Textile Industry of Westmünsterland], Dortmund: Dortmunder Vertrieb für Bau- und Planungsliteratur.

Hodson, M. (2008), 'Old industrial regions, technology, and innovation: tensions of obduracy and transformation', *Environment and Planning A*, **40**, 1057–75.

Hudson, R. (1994), 'Institutional change; cultural transformation, and economic regeneration: myths and realities from Europe's old industrial areas', in A. Amin and N. Thrift (eds), *Globalization, Institutions, and Regional Development in Europe*, Oxford: Oxford University Press, pp. 196–216.

Hudson, R. (2005), 'Rethinking change in old industrial regions: reflecting on the experiences of North East England', *Environment and Planning A*, **37**, 581–96.

Hudson, R. and D. Sadler (2004), 'Contesting works closures in Western Europe's old industrial regions: defending place or betraying class?', in T.J. Barnes, J. Peck, E. Sheppard and A. Tickell, (eds), *Reading Economic Geography*, Malden, Oxford & Carlton: Blackwell Publishing, pp. 290–303.

Lam, A. (2002), 'Alternative societal models of learning and innovation in the knowledge economy', *International Social Science Journal*. **171**, 67–82.

Lee, C.-W., Y.-C. Park and E.J. Kim (2000), 'The restructuring of textile and apparel industry and the spatial division of labour on production system in Daegu', *Journal of the Korean Geographical Society*, **35**, 207–25 (in Korean).

Lorenzen, M. (2005), 'Editorial: Why do clusters change?', *European Urban and Regional Studies*, **12**, 203–08.

McGillivray, F. (2004), *Privileging Industry: The Comparative Politics of Trade and Industrial Policy*, Princeton, NJ: Princeton University Press.

Martin, R. and P. Sunley (2006), 'Path dependence and regional economic evolution', *Journal of Economic Geography*, **6**, 395–437.

Maskell, P. and A. Malmberg (1999), 'Localised learning and industrial competitiveness', *Cambridge Journal of Economics*, **23**, 167–85.

Melachroinos, K.A. and N. Spence (2001), 'Conceptualising sunk costs in economic geography: cost recovery and the fluctuating value of fixed capital', *Progress in Human Geography*, **25**, 347–64.

Moon, C.I. and S.-Y. Rhyu (2000), 'The state, structural rigidity, and the end of Asian capitalism: a comparative study of Japan and South Korea', in R. Robison, M. Beeson, K. Jayasuriya and H.-R. Kim (eds), *Politics and Markets in the Wake of the Asian Crisis*, London & New York: Routledge, pp. 77–98.

Morgan, K. (1997), 'The learning region: institutions, innovation and regional renewal', *Regional Studies*, **31**, 491–503.

Morgan, K. and C. Nauwelaers (1999), 'A regional perspective on innovation: from theory to strategy', in K. Morgan and C. Nauwelaers (eds), *Regional Innovation Strategies: The Challenge for Less-Favoured Regions*, London: The Stationery Office and Regional Studies Association, pp. 1–18.

Myrdal, G. (1957), *Economic Theory and Under-Developed Regions*, London: Gerald Duckworth & Co. Ltd.

OECD (2001), *Cities and Regions in the New Learning Economy*, Paris: OECD.

Park, K.-T. (1997), 'South Korean economic development and the concentration of the textile industry in Taegu since the 1960s', *Journal of the Korean Geographical Society*, **32**, 359–70.

Reckfort, J. and M. Ridder (1996), *Die münsterländische Textilwirtschaft: Bedeutung, Struktur, Entwicklung und Potentiale zur Stärkung der Wettbewerbsfähigkeit* [The Münsterland Textile Economy: Importance, Structure, Development and Potential to Strengthen Competitiveness], Münster: Schriften zur Textilwirtschaft, Bd. 48.

Röller, L.-H. and C. von Hirschhausen (1996), 'State aid, industrial restructuring and privatization in the new German Länder: competition policy with case studies of the shipbuilding and synthetic fibres industries', WZB Social Science Research Center Berlin, Discussion Paper FS IV 96–13.
Saxenian, A. (1994), *Regional Advantage: Culture and Competition in Silicon Valley and Route 128*, Cambridge, MA & London: Harvard University Press.
Schamp, E.W. (2000), *Vernetzte Produktion: Industriegeographie aus institutioneller Perspektive* [Networked Production: Industrial Geography from an Institutional Perspective], Darmstadt: Wissenschaftliche Buchgesellschaft.
Schamp. E. W. (2005), 'Decline of the district, renewal of firms: an evolutionary approach to footwear production in the Pirmasens area, Germany', *Environment and Planning A*, **37**, 617–34.
Shapira, P. and G. Fuchs (eds) (2005), *Rethinking Regional Innovation and Change: Path Dependency or Regional Breakthrough*, New York: Springer.
Sohn, T.-S. (1999), '"Milan of Asia" project for Taegu hits early snag', *Korea Herald*, 1 March.
Stopford, M. (1997), *Maritime Economics*, London & New York: Routledge.
Storper, M. (1997), *The Regional World: Territorial Development in a Global Economy*, New York & London: The Guilford Press.
Stråth, B. (1986), 'Redundancy and solidarity: tripartite politics and the contraction of the West European shipbuilding industry', *Cambridge Journal of Economics*, **10**, 147–63.
Todd, D. (1991), *Industrial Dislocation: The Case of Global Shipbuilding*, London & New York: Routledge.
Tödtling, F. and M. Trippl (2004), 'Like phoenix from the ashes? The renewal of clusters in old industrial areas', *Urban Studies*, **41**, 1175–95.
Toyne, B., J.S. Arpan, A.H. Barnett, D.A. Ricks and T.A. Shimp (1984), *The Global Textile Industry*, London: George Allen & Unwin.
Van Grunsven, L. and F. Smakman (2005), 'Industrial restructuring and early industry pathways in the Asian first-generation NICs: the Singapore garment industry', *Environment and Planning A*, **37**, 657–80.
Whitley, R.D. (2000), 'The institutional structuring of innovation strategies: business systems, firm types and patterns of technical change in different market economies', *Organization Studies*, **21**, 855–86.
Woo, Y.-S. (2003), 'Spatial characteristics of production networks in the shipbuilding industry', *The Journal of Korean Economic Geography*, **6**, 99–117 (in Korean).

PART 5

STRUCTURAL CHANGE, AGGLOMERATION EXTERNALITIES AND REGIONAL BRANCHING

22 The evolution of spatial patterns over long time-horizons: the relation with technology and economic development

Jan Lambooy

1. Introduction

The relation between technology and economic development has been investigated extensively. Their combined impact on spatial development, in particular on urbanisation, the location of firms and the emergence of new regional economic patterns, still needs to get more attention. With *spatial patterns* (or spatial structures) we do not only mean the way in which physical objects are structured in space, but also how social and economic activities and actors are situated, and how they behave and interact in processes like relocation, urbanisation and regional economic production. Space can also be interpreted as Perroux (1950) did, first as a set of *locations* in the *geographical* or *physical space*, and second as a set of *network relations* in the *topological space*, where geographical relations are not important, only their relative locations are. In recent times the concept of 'networks' is used to indicate relations of various kinds and with various perspectives. It can be used for physical networks of infrastructure, for social relations and as an organisational concept for the coordination of a set of units that are part of a firm or other kinds of organisation (Huber, 2007). Many relations are enabled by physical structures, such as road infrastructure, and telecommunication, by the internet. Spatial patterns reflect technology and economic development in various ways, however, often with a certain time lag. Spatial development evolves more gradually, over longer and different time-paths, than the other two elements of this 'triangle'. Technology has an impact on economic development and the use of space, but there is a reverse impact also. Economic growth and new technologies, by investments in capital goods, in R&D and knowledge, lead to continuous changes of the spatial and economic structures. This is often related with the differential effects of the rise of productivity (Maddison, 1991; Pasinetti, 1981). In certain favourable urban regions, innovations, technology and economic growth have better prospects than elsewhere (Jacobs, 1984; Krugman, 1995). As Martin (1999, p. 76) contends: 'While spatial agglomeration is a key feature of the contemporary economic landscape, so is structural change. The question of uneven regional development is not a static one, but is continually evolving, entails major qualitative and quantitative change'. Urban development is also related with certain kinds of infrastructure and technology, like airports, rail infrastructure and, more recently, with the development of networks for the transmission of electricity and electronic data (Peschel, 1989). Urban development in the last centuries strongly reflects the increased speed of the changes in the technological and economic basis of our society.

Evolutionary economic geography (EEG) can contribute to the investigation of these processes of co-evolution in at least four ways. First, it acknowledges space and time as important factors in economic development. Second, it pays attention to continuous differentiation and selection, and to changing technology and economic structures. Third, EEG attempts to analyse the changing composition of firms and sectors, influenced by various innovations of scientific, technological, institutional and organisational nature. Fourth, and finally, it emphasises the continuous process of adaptive behaviour of economic actors, like firms, when the 'environment' changes, for instance when technology, markets, locations and institutions change. For EEG it is important to consider the question of how firms, sectors, cities and regions adapt to new opportunities and barriers to growth. Long-term changes in spatial patterns can be related to pervasive technologies and to the flexibility of economic systems to adapt to the new challenges.

Evolutionary approaches are not restricted to gradual changes, they can also cover longer time-horizons and structural changes of economic, technological and spatial patterns (Boschma, 1994; Lambooy, 1984; North, 1990, 1994; van der Knaap, 1978; van Duijn and Lambooy, 1982; Wenting and Boschma, 2006).

This chapter is necessarily selective, and it is not the place to discuss in depth the theoretical differences among authors who investigate nature of the influence of scientific development on technological development. Some authors emphasise the gradual development, whereas others emphasise the influence of revolutionary structural changes over longer time-horizons. We refer to Anderson et al. (1988), Cohen (1994, 2007), Mokyr (2002) and Hellyer (2003) where this issue is discussed, in particular with regard to the influence of the scientific revolution on technology and economic development. However, here the focus is on the views of the interrelations of waves of new technology on long-term economic development and on spatial patterns.

In section 4 the issue of the development patterns of the economy (endogenous, gradual, cyclical, or exogenous and in phases) is explored with the theories of 'long waves' (Kondratieff, 1926; van Duijn, 1983) and of 'phases in the development' (Maddison, 1991).

In this chapter three concepts are used to investigate the various phenomena associated with the developments of technology, economy and spatial structures. The first one is the concept of *general purpose technology* (GPT). This can be used to investigate the development of pervasive technologies. A GPT is a set of technologies that is pervasive to all economic production and consumption. Lipsey et al. (2005, pp. 12–13) describe GPTs as follows:

> They begin as fairly crude technologies with a limited number of uses and they evolve into much more complex technologies with dramatic increases in the range of their use across the economy and in the range of economic outputs that they help to produce. As they diffuse through the economy, their efficiency is steadily improved. As mature technologies, they are used for a number of different purposes, and have many complementarities in the sense of cooperating with many other technologies.

As such, a GPT can be seen as a basis for new waves of improving productivity (Dosi, 1982; Dosi et al., 1988; van Duijn and Lambooy, 1982). A GPT can be investigated in various ways. For our purpose, a selection is made on the basis of the impact on spatial patterns. It can be argued that the impacts of GPTs have been extremely important for both space and economic development. Our society and our spatial configurations have been completely altered by the various GPTs, more in particular by those that were the basis of the Industrial Revolution in the nineteenth century. The most recent wave of new technologies, connected with the introduction and dissemination of ICT (information and communication technology) since the 1950s, will also have its effect on spatial configurations in physical space, but it is not yet entirely clear how this will work out in spatial configurations. In section 5 we return to this issue.

The second concept used is that of a *long wave of economic development*, which denotes the long-term economic development and the changing resource use and the composition of production (Maddison, 1991; van Duijn, 1983; van Duijn and Lambooy, 1982). An evolutionary approach seems to be useful to examine the co-evolution, or the integrated view of the development of the 'triangle' of relations. Nelson (2005, p. 8) contends: 'A viable economic growth theory must recognize the evolutionary nature of processes of technological advance, and of the ways the structure of inputs, outputs and institutions is molded and molds the advance of technology.'

The third concept used is that of a *complex adaptive system*. This concept is based in an evolutionary approach and is useful to investigate more complex structural changes (Anderson et al., 1988; Antonelli, 2008; Arthur, 1999; Hayek, 1937 and 1945, cited in Caldwell, 2004). Metcalfe et al. (2006) use this concept to examine the nature of economic growth. They contend that capitalism is restless and adaptive. The basis for growth is the process of 'self-transformation, the emergence of macro-structure from micro-diversity, and coordination through market processes' (Metcalfe et al., 2006, p. 7). They also argue that it is the generation and resolution of economic diversity that is the principal source of growth.

The main concern of this chapter is to explore the interrelations between long-term economic, technical and spatial development, which are dealt with more specifically in sections 4 and 5. However, we first pay attention to the approach of evolutionary economic geography in section 2. Then the following section 3 investigates the interrelationship between technology and the development of spatial structures (urbanisation and interregional economic differentiation) in complex adaptive systems. In sections 4 and 5 the relations of technology, economic development and spatial patterns over long time-horizons are examined with the theories of the long-wave and the development of phases. In section 5 special attention is given to the influence on urban and regional location patterns of the more recently developed new GPT – information and computer technology (ICT). This chapter is then finalised by drawing some conclusions based on the previous sections.

2. Evolutionary economic geography, spatial structure and time

The evolutionary approach to economic geography (EEG) differs in various dimensions from that of the approaches of the neoclassical mainstream and new economic geography (NEG). This chapter will not explore this, because others, such as Martin (1999) and Boschma and Frenken (2006), have extensively developed that issue. Here the main focus is the long-term changes in the relations in the 'triangle' of technology, space and economic development. This interrelation can be investigated by using the viewpoints of *evolutionary economics*. This theory focuses on the continuous emergence of novelties and selection processes with unpredictable outcomes (Arthur, 1994, 1999; Boschma and Lambooy, 1999; Nelson and Winter, 1977). It is a long-term emergent process, caused by chance and exogenous events but also by endogenous self-reinforcing processes with an interactive feedback nature, like with the transfer of knowledge between firms and other component parts of the economic system (Lambooy, 2002, 2006; Leidig, 2002). New spatial patterns emerged during the Industrial Revolution, more in particular the industrialised cities. More recently in the time of the ICT revolution, Silicon Valley is an example of the result of the combined influences of technology, economic development and new forms of organisational structures. Urban regions, firms and consumers are again in a period of restructuring their space, adjusting to new technologies. In many regions old technologies and old sectors have remained important. It is difficult for economic actors to alter their perspective from past successes towards new opportunities, a situation that is sometimes indicated by the concepts of 'path dependency' and 'lock-in'. This is one of the reasons why old manufacturing regions have great difficulties in using new technologies and developing new kinds of spatial structure. They need to develop a strategy of 'path-creation' into new fields of opportunities (Antonelli, 2008; Boschma and Lambooy, 2000; David, 1985; Martin, 2006; Martin and Sunley, 2006). However, this can take a long time to realise. Technological, economic and spatial restructuring processes use different time-scales.

Arthur contends: 'Economic patterns are ever changing, showing perpetually novel behaviour and emergent phenomena. Complexity portrays the economy, not as deterministic, predictable, and mechanistic, but as process dependent, organic, and evolving' (Arthur, 1999; p. 107). It is not entirely clear how the processes of economic and technological changes affect spatial structures and vice versa (Antonelli, 2003, 2008; Lambooy, 2002; Martin and Sunley, 2006). However, with the concept of 'complex adaptive behaviour' it can be shown that evolutionary economics not only emphasises novelty, selection, retention and diffusion, but also the continuous adaptive behaviour of the actors to survive. The survivors are not primarily those who are the 'fittest', but those who best adapt to new circumstances in their economic, spatial and institutional environment. The topic of change and adaptive behaviour belongs to the core of evolutionary and institutional economics. For EEG the development of spatial patterns can be added to that approach (Boschma and Lambooy, 1999).

In EEG, several dimensions of technology are distinguished, in particular the differentiating effects of technology, heterogeneity of economic structures and the change of structures over time (Geels, 2002; Pasinetti, 1981; Peneder, 2002). Evolutionary economics accepts heterogeneity and 'real dynamics' (in contrast to the 'comparative statics approach'), the change of structures by the introduction of new technologies, with concomitant unpredictable outcomes (Boschma, 2004; Doi and Mino, 2005; Lambooy, 2002; Lambooy and van Oort, 2005; Metcalfe, 1994). Metcalfe discusses the specific value of an evolutionary approach. In his words 'adaptive systems' with various emergent and unpredictable processes are inherent properties of evolutionary development. (Metcalfe, 1994, p. 933) contends: 'A distinctive feature of the evolutionary approach is its adoption of a behavioural theory of the firm and its focus upon learning processes and adaptive behaviour'. Adjustment is extremely important in the case of the introduction of pervasive technologies (GPTs). For our purpose, we emphasise that various technologies can converge and develop into general purpose technologies (GPTs). In the words of Metcalfe et al. (2006, p. 8):

> All the aggregates we deal with are emergent consequences of the interaction between different industries in relation with the growth of productivity and the distribution of the ensuing incre-

> ments in demand. [. . .] The productivity growth in one sector spreads to others via income and expenditure flows through markets.

The opposite case of stagnant sectors and regions is also important to investigate the time needed for adjustment to new conditions and for restructuring. This does not only need new kinds of entrepreneurs and technologies, but also new attitudes.

EEG uses micro and meso levels of analysis, much less the macro dimension. Lewis (2004; p. xvi) argues that, to explain growth, one needs to turn to the micro and the meso levels. Firms and consumers are often seen as the basic units of analysis at the *micro* level. At the *meso* level, firms, industries and sectors, but also regions and cities can be taken as a central focus for investigation (Lambooy, 1989). At the *macro* level, regions and countries can be investigated, but mainly as a result of interactive dynamic relations at the micro and meso levels. Macroeconomic development is often measured with indicators like GDP and employment. Macro structures, once developed, influence or condition human behaviour. In micro-analysis the emphasis is on actors and behaviour. Actors are involved in emergent processes, in unpredictable ways, with uncertain macroeconomic and spatial outcomes. They adjust continuously to changes in their environment. The relation between microeconomic actions and macroeconomic outcomes is difficult, more in particular over longer time-horizons. Although the explanation of novelties has to be connected to micro behaviour, in our investigation mainly the outcomes are considered.

The concept of 'the economic system as a complex adaptive system' and theories of 'long waves' and 'development with phases' are interesting attempts to understand the nature of these longer time-horizons. Scott (2006) argues that the investigation of technology and production systems is decisive to understand the development of regional and urban economies.

Space is an integral part of human societies and evolves together with human activities. It can be conceived as an integration of human activities, their networks and their various environments as an adaptive complex system of 'human ecology' (Lambooy, 1984, 1989). Actors adjust to changing activities and conditions in an emergent process. Urbanisation processes and interregional differences in economic growth reflect the forces of technology and broader human processes, like institutional forces.

In the next section we further explore the interrelations between technology and spatial development.

3. Technology and space

Technology can be defined in various ways. The most usual view in economics is to consider technology as a bundle of human capabilities, skills and routines that is used in production, primarily the production of goods but also the production of services (like surgery, or truck-driving). Technology is used to transform goods, to enable a certain kind and quality of services, and to enable the use of goods in consumption. In this sense of the use of material goods, knowledge and skills, technology is important in consumption too, like in the use of computers and TVs. Technology can also be conceived as knowledge embodied in the 'hardware' (or 'tools') of capital and consumption goods. It primarily consists of knowledge as an attribute of actors, combined with material aspects in tools, equipment, buildings and infrastructure. The effects of technology on productivity and incomes are undeniable, more in particular when GPTs are widely used in the economy. Successful innovation requires its integration into the social and organisational contexts (Malerba, 2006).

Lipsey et al. (2005) also include *organisation* in the concept of technology, because that is a necessary condition to enable the process of production and consumption. Organisation can be a condition for the development of technology, but it can also be a result of the introduction of new technology (Kim, 2005). Chandler (1962), Chandler and Mazlish (2005) and Gottmann (1961) have argued that the rise of skyscrapers, the technology of steel construction, the introduction of the elevator, together with telecommunication and the car, have forced corporate America to adapt both their economic and their spatial organisation. Currently, the introduction of ICT again forces many enterprises to adapt to new challenges and to reorganise their production and benefit from 'new windows of locational opportunities'. This process of reorganisation in relation to the rise of new technologies has had a strong impact on the organisation of economic space, in the sense of locations, as in that of the structure of connections and networks (Gay and Dousset, 2005). In section 5 ICT is dealt with more extensively.

Andersson (1981) has indicated three fundamental factors determining structural economic change: (1) reallocation of resources; (2) quantitative expansion of the resource base; and (3) technological development. In earlier theories of economic growth the first two factors were considered to be more important for regional economic development. The focus of economic theories has changed to the third factor: technology and other forms of knowledge.

Technological development has three principal influences on economic and spatial structures:

1. *Rising productivity*, with concomitantly increasing incomes and consumption opportunities (houses, cars, household appliances). It also leads to important shifts in the configuration of manufacturing industries and services. Combined, these shifts have a strong impact on the location of firms and households.
2. *Declining costs of distance*. The declining costs of transport and communication of goods, people and information can restructure the space of locations of production and consumption. However, to realise these cost-reductions enormous investments are needed, which often have been financed with public means.
3. *Stronger agglomeration advantages*. The combination of rising incomes and lower transport costs with the availability of trains and cars has led to spill-over effects and strong urbanisation processes with the concentration of innovation and economic growth in certain regions.

Within agglomerations, the economic actors can use space and information more efficiently, which can result in lower transport and transaction costs as well as in a more efficient organisation of production and consumption.

The development of the new GPTs has, directly and indirectly, affected the various spatial patterns, as Boschma (1994) has investigated for Great Britain and Belgium. Here again it is necessary to emphasise that the time-horizons of the gradual and structural changes within the triangle of relations can differ. The introduction of technologies like steam power, electricity, engines for cars and airplanes, and more recently information and communication technology (ICT), has provided important drivers of changes in economic and spatial structures. However not all regions and cities have benefited equally and in the same period of time from this development. There are also differences in opinion about the mechanisms of development. Rosenberg and Trajtenberg (2004) believe that the change from the old source of energy for manufacturing (water power and mills) to a new GPT (steam power and the steam engine) was essential for the urbanisation processes and economic growth. Kim (2005) argues that the intermediate factor was not the shift from water power to steam, but the changing *organisation* of factories, with an intense use of labour, which made it necessary to be located in urban regions. Nevertheless, the steam engine was essential for the rise in the productivity level and enabled the rise of new locations for manufacturing industries. When various kinds of technology converge (like electricity and steel construction technology for high-rise buildings), the economic, organisational and spatial effects can be important for the entire society. With regard to the effects on spatial structure, a principal effect is the decline in distance costs and distance barriers, which can lead to more productive spatial economic configurations, but also to the ease of communication in networks, with clustering and agglomeration advantages (Boschma and Kloosterman, 2005). As Jane Jacobs (1969, 1984) had already observed, the diffusion of technological knowledge is easier in differentiated agglomerations than in regions where the entire economy depends on a restricted number of economic activities. This is one of the reasons why large urban regions with a varied economic structure can be considered as strong 'adaptive systems' and as 'source regions' for new ideas (Lambooy, 2006). According to modern theory, these agglomerations have growth effects on the entire society (Krugman, 1995; van Oort, 2003). New knowledge, like technology, and the mechanisms and channels of its diffusion (externalities, spill-overs, networks), are key issues for the understanding of innovations and the development of the economy and of spatial patterns (Castellacci, 2007).

In the following sections the relation between technology and economic growth over longer time-periods is considered. The approaches focus primarily on the Industrial Revolution in the nineteenth century, but more recently much interest has been raised by the possible important spatial impacts of the ICT revolution.

4. Economic development over long time-horizons

In macroeconomic approaches, economic growth is investigated by looking at the development of GDP and per-capita GDP, often without a perspective of longer periods of time. Well known is the statement of Keynes, 'In the long run we are all dead', but he forget to mention that economic and spatial structures often affect behaviour over long periods of time. The analysis of technological changes and the structural changes of the economy, like that of the sectoral composition and the altering significance of natural and human resources over longer time-horizons, is very important to understand the drivers of economic growth. It is also necessary to look at the behaviour of the micro and meso levels of the economy. The roles of entrepreneurs, institutions and agglomeration forces are also missing in many macroeconomic studies, with a concomitant lack of understanding of the dynamic forces behind development. However, in more recent approaches on economic growth, technology and entrepreneurs are increasingly considered as a main cause for economic growth, whereas institutions (like taxes, social security, labour market policy, etc.) can hamper or enable growth (Helpman, 2004; Nelson, 2005; Solow, 2003).

When examining economic growth over a long period, starting with the time of the Industrial Revolution until about 1870, it can be observed that new technologies and the consequent rise in the productivity level are at the basis of important structural changes (Maddison, 1991; Pasinetti, 1981). New technologies and organisational production methods increased productivity and caused millions of people to shift to the strongly developing manufacturing sectors, with a strong effect on urban growth. The impact of the Industrial Revolution on the economic and spatial structure was enormous. Many new regional and urban developments could be observed. The rise of 'Black Cities' in England, Germany and the US are proofs of this impact. Boschma (1994) has investigated the development of interregional shifts of economic activities during the period of the Industrial Revolution in Great Britain and Belgium.

Cohen (1994) and Hellyer (2003) have contended that the relation with the 'Scientific Revolution' has been decisive for the development of technology and the Industrial Revolution. Mokyr (2002) emphasised the endogenous nature of this process of economic growth by looking at the strong interrelations between applicable scientific research and technological and industrial developments. He found that this 'application-oriented approach to science' was particularly strong in the Anglo-Saxon world, a main reason for the Industrial Revolution to happen there first and only later in the continental countries. According to the new endogenous growth theory, R&D and investment are endogenous to the economic process (Helpman, 2004; Romer, 1986). But that theory does not encompass the element of time and the sources of dynamism. The discussion on endogenous or exogenous influences has also been dealt with in the 'long wave' theories.

Many theories have been developed on long-term economic growth in relatively ordered patterns (van Duijn, 1979, 1983). In Russia Kondratieff (1926) developed his 'theory of economic cycles', that lasted 50 to 60 years. The Dutch van Gelderen (1913, under the pseudonym Fedder), the Belgian Mandel (1975), and the Austrian Schumpeter (1912 [1934]) attempted to construct theories explaining the 'cycles', or the 'long waves' of economic development. They emphasised the existence of a regularity of the waves, meaning that certain patterns of the growth process return in every new cycle in ordered patterns. The development and the demise of cycles were inherently endogenously determined, although every new cycle was based on new technologies and new impulses from investment in capital goods and infrastructure. The (young) Schumpeter emphasised the role of entrepreneurs (Schumpeter, 1912 [1934]). He was not so explicit on the longer cycles, as he focused more on business cycles with shorter amplitudes than the 50 or 60 years of Kondratieff. He explained the rise of new growth, the crisis and the recovery by the introduction of fundamental innovations by entrepreneurs and the replacement of old by new products. His explanation was that crises were a motivating factor for entrepreneurs to be risk-taking and innovative. He also contended that economic crises could be explained endogenously, as a result of investments in R&D and of exploring new opportunities and new sources for profit enabled by the introduction of innovations. Later Schumpeter (1943) emphasised the role of larger corporations in the innovation process, because the costs of R&D became too high for the smaller firms. Nevertheless he accepted that even then economic depressions were challenges to entrepreneurs, and could stimulate research and risk-taking, the rise of new technologies, new firms and new sectors, but could also lead to 'creative destruction', the demise of old firms and sectors, with possible new periods of crisis (Kleinknecht, 1987). Also Mensch (1975), the German researcher of innovations and long time-horizons, emphasised the rise of innovations in times of crisis. He contended that the scientific and technological knowledge (inventions) developed more evenly in time than innovations (introduction in production and markets). Innovations were peaking in periods of economic crisis.

The changing economic structure, resulting from the rise of new firms and new products and technologies, was very important to understand later bursts of rising productivity and economic growth. Kondratieff, van Gelderen and Mandel emphasised the impact of the cycle of investments in capital goods and infrastructure as an endogenous process in economic development, although they measured the cycles mainly in units as GDP, employment and prices and not in the nature of technology, goods or infrastructure. After a certain period of time, new investments were necessary and a new cycle started all over again.

Maddison (1991) does not start with the assumption of regularity and of endogenous processes. According to him, technological changes cause fundamental adjustments of economic structures, but he refuses to accept the regularities of waves insisted on by other authors. Maddison accepts not only technology but also institutional forces (government policies) and exogenous shocks as causes for periods of 'breaking points'. He contends (Maddison, 1991, p. 149):

> Structural changes reflect two basic forces that have operated on all countries as they reached successively higher levels of real income and productivity. The first of these is the elasticity of demand . . . and the second basic force has been the differential pace of technological advance between sectors.

With Rostow (1978), he speaks of 'stages' or 'phases', not 'cycles' or 'waves'. Breaking points are the results of exogenous incidents occurring without regularity. But, within the phases, coherent development can be shown, which can be seen as a process of endogenous complex adjustment. Maddison (1991, p. 112) contends:

> In the 170 years since 1820 one can identify separate phases which have meaningful internal coherence in spite of wide variations in individual country performance within each of them. Phases are identified, in the first instance, by inductive analysis and iterative inspection of empirically measured characteristics.

He does not focus on regularities that the 'wave theorists' adhere to. Instead, he supports a more neutral view (Maddison, 1991, pp. 111–12). He concludes (1991, pp. 123–24): 'The move from one phase to another has been caused by system shocks (. . .) usually governed by exogenous or accidental events which are not predictable'. Instead of a regular and predictable 'cycle', he accepts shocks or strong interruptions of a certain growth-path, causing strong changes in the economic structure. He does not deny the endogenous character of development *within* the phases.

In my opinion Maddison seems to be closer to empirical validation than the other views as far as the effort to find periods of growth is concerned. His analysis is focused on macroeconomic and sectoral indicators of economic development. However, he does not pay attention to the role of entrepreneurs and their efforts of innovation, nor to the evolutionary processes of the creation of new variety, selection and the development of new kinds of urban and regional economies. His views and those of the Schumpeterian approach can be regarded as being complementary. Aghion and Howitt (1998, p. 244) contended that it is possible to combine the endogenous growth theory with the Schumpeterian approach: 'Cyclical downturns may be the price that society needs to pay in order to complement the GPTs that deliver the long-run growth.'

The result of the process of unpredictable innovation dynamics can sometimes completely alter economic structures and locations. During the nineteenth and twentieth centuries, the principal structural change in the production structure and the spatial structures of the nineteenth century has been the almost tragic decrease in agricultural employment, combined with a strong migration to urban regions. This stream of migrants was within the home countries and outward to the New World. Today's modern world still has a strong and productive agricultural sector, be it with only about 2 to 4 per cent of employment. At the same time, the manufacturing sector increased until the early 1960s, after which employment and the share of GDP have declined to percentages of between 10 and 20. The various service sectors are now as important as the agricultural sectors used to be in 1800. Their contribution to employment and GDP has risen to between 65 per cent and 80 per cent. This structural change came about very gradually, but the main cause was undeniably the changing productivity and the differences between sectors (Nelson and Winter, 1977; Pasinetti, 1981; Peneder, 2002). At the global level many of these processes continue, related to the 'global shift' to Asian countries.

Technological development came in stages, economic development showed breaking points, but the long-term tendency of rising productivity and changes in sectoral composition has been the same from 1800 on. Van Duijn (1983) has raised the idea that the nature and the location of economic growth will be influenced by the rise of new global economic centres, as the USA and Japan showed in the twentieth century, and – as we now can add – China and India in the twenty-first century. In the present time we observe the results of this structural change, in the altered structures of demographics, space and global connections. The impacts of structural changes on the urbanisation structures have been enormous. The Industrial Revolution caused a strong concentration of employment and population in urban areas. Both processes (industrialisation and urbanisation) needed huge complementary investments in infrastructure. The impact on space and on the location of economic activities had dramatic effects on the lives of millions of people.

To sum up: both approaches, the cyclic or long-wave theory and the theory of phases support the idea that technological developments, economic growth patterns and spatial changes show strong interrelations, when observed over long periods. More in particular the physical space has undergone major changes by the huge urban developments. Organisations and the nature of social and economic networks did also change. In more recent times the spatial patterns have been changed by the introduction of the 'computer and network technology'; however, the change of physical space seems unlikely to develop at the scale seen in the nineteenth century. In the next section some of the recent developments are investigated.

5. The impact on space of ICT as a GPT

During the last three decades, the influence of ICT on society has become an important issue (Giaoutzi, 1989; Lambooy, 1987). The effects that received most attention were the impact of this new GPT on productivity levels and the changing nature of communication and organisation. The use of computers in business and households disseminated quickly. In the second half of the 1990s, the rise of the internet stimulated a further growth of the use of ICT. For instance, in 2003, 80 per cent of the Dutch population had access to a computer and more than 90 per cent of all businesses owned one or more computers. Furthermore, an increasing part of the economy relies on the support of ICT, which is illustrated by the more than 60 per cent of all people whose daily work involves working at a computer (Statistics Netherlands, CBS, 2003). A similar process of diffusion can be observed in other countries. The economic effects seem to be clear as most economists sustain the idea that productivity increased, and that the opportunities for the communication and the optimisation of logistic and other networks were considerably improved. ICT has also had an effect in terms of its influence on the organisational structure of companies and markets, particularly related to its impact on transaction costs.

The spatial effects can be distinguished in those on the global level and the impact on urban regional space. Like Kim (2005) observed about the spatial impact of the introduction of the steam engine, it is not the PC as such or 'teleworking' that is the main driver of change, but the organisational restructuring of corporations and of the global commodity chains. In the case of ICT, the evolution of new spatial patterns may seem different from those in the time of the Industrial Revolution. The changes will develop less strongly in physical space, as in the sense of the rise of new kinds of city, but more in the organisational structure and the relations in networks. Many changes can be observed at the global level of the division of labour. New 'windows of new locational opportunities' are created in the 'emerging economies' (Boschma, 1994). ICT has created in Taiwan, Israel and India new opportunities for 'mobile innovators' (or 'New Argonauts'), living and working in two economic environments, the USA and their home regions (Saxenian, 2006). This new GPT has not only been the source of the rise of Silicon Valley as a new agglomeration in the San Francisco Bay Area, but has also offered new opportunities to multinational enterprises and innovative small firms. This new mobility has raised a new and rapidly growing number of publications on the nature of communication and the rise of networks and global networked organisations (Huber, 2007; Lambooy, 2006). In spatial terms, it could lead to a scenario in which every workplace with access to the worldwide virtual network is a viable alternative to urban office concentrations as a place in which to work effectively and efficiently. Companies are thought to become 'footloose' because ICT renders them less dependent on the local market. For the regions and the regional real estate market, this will arguably have significant consequences. For many companies, the traditional urban setting might no longer be the automatic choice of location, although empirical evidence shows that they still prefer locations within metropolitan areas (Weterings, 2006).

The effects on urban space are as yet not entirely clear. Cairncross (2001) coined the term *death of distance*. *The Economist* went on to describe this 'end of geography' as the most significant force to shape society over the next 50 years. Its impact will be comparable to the introduction of electricity. The rapid developments and the widespread adoption of ICT have led to many speculations about the spatial effects of this technology. Sometimes teleworking and the use of the internet are seen as a possible future replacement of the office, leading to further urban sprawl. Households and companies seem to be, to some degree, resistant to these forms of communication 'without a face'. People can communicate through e-mail, but their most important decisions are still made during face-to-face contacts. In fact, the increase in appointments made by e-mail has actually intensified formal and informal meetings between people (Storper and Venables, 2002). Of course, at the same time the networks became worldwide as well. Working at home is possible, but remains up to now often confined to more standardised administrative work or very specialised personalised work (for example editing, translating and writing), which requires no resources other than knowledge, a PC and a broadband connection. Yet, even in these activities, people seek contact with their colleagues in the workplace or at clubs and social events. The work conducted and the relations, in a setting other than at home, form an essential component of modern lifestyles and routines. However, spatial proximity can be, at least partly, replaced by activities and relations in topological 'economic space' or networks (Boschma, 2005; Huber, 2007).

Although the effect of ICT on spatial development seems to be more limited than the earlier prophets predicted, there still are some important effects, for instance on the business organisation, on the global division of labour, on shopping behaviour and on the ways of intercommunication of many households. The processes of suburbanisation and polycentricity have been established in many countries for some time now, but most studies to date show that ICT is playing a facilitating role in these processes. The effect of ICT on the development of other kinds of urbanisation may become stronger in the future (Graham and Marvin, 2001; Spufford, 2006).

Bartley contends that the effects of ICT on urban de-concentration are less certain than was expected by adherents to the 'death of distance' hypothesis (Bartley, 2006). However, Sassen (1991) argued that three transformations can be discerned: (1) the geographic dispersal of manufacturing and the decline of industrial centres; (2) the growth of the financial services; and (3) the changing relations between cities, more in particular between the global cities. According to Scott (2006), contrary to what many believe, larger urban regions remain attractive in terms of the composition of economic activities rather than the so-called 'quality-of-life indicators' or the impact of ICT. It seems, however, that workers do not just act with the opportunities of ICT in mind. They still prefer urban locations with a better quality of life for knowledge workers, as Florida (2002) and Glaeser and Gottlieb (2006) contend.

Our conclusion is that the resistance to spatial change presumably is now stronger than at the beginning of the earlier revolutionary waves of economic and technological development, because we already live in a strongly urbanised world.

New conceptual frameworks seem to be necessary to effectively analyse the spatial developments of path-breaking new technologies like ICT. Until now, different terms have been used such as 'polycentric urban structures', 'networked cities', 'externalities', 'spaces of flows', 'penetration of computer automation', 'splintering urbanism' and 'dispersed de-concentration'. These are the terms that can form part of the major structuring elements of future research into the relationship between ICT on the one hand and spatial (urban) development on the other. Because these concepts were introduced only rather recently, there is as yet little research available to confirm or disprove their validity. In order to judge the impacts of ICT a longer time-horizon is needed. Previous technological developments fostered a strong process of urbanisation, but that took a long time to evolve. It seems that comparable developments of the changes in technology, urban development and sectoral patterns can be observed in present-day East Asia (Gill and Kharas, 2009).

6. Some conclusions

In this chapter the development of the interrelations between technology, economic development and spatial structures over a long horizon have been explored. Three concepts were used to investigate the development of this 'triangle'. First, *GPT*, the pervasive technology that revolutionised many production systems, increasing productivity and altered the impact of distance on human contacts. The second concept was the idea of *long waves of economic development*, with the possibility that these could either have a cyclical and endogenous character, or that 'phases' dominated with exogenous influences in critical periods, but inside the phases could develop endogenously. The third concept, originating from evolutionary theory, the *complex adaptive system*, can be used to explain the continuous rise of unexpected novelties, in combination with the capacity of the human actors and firms to adjust to the new environment of altered markets, institutions and spatial patterns. New kinds of organisation, network and location are the result of this continuous process of adaptation. The technological development caused a shift of the composition of sectors, resulting in the strong rise of productivity, first in agriculture, then in manufacturing. This resulted in a dramatic decline of agricultural labour in combination and a strong rise in manufacturing labour and the rise of industrial cities. Later, this shift went further with the increasing productivity in manufacturing firms, and a growing dominance of the service sectors in employment and in GDP. The urban regions saw an increased attraction and increased in population and jobs. During the Industrial Revolution a major process of changes in the triangle occurred. Urbanisation was the clear spatial result of that revolution. In the more recent period a new major GPT is developing, ICT, but the spatial effects are not entirely clear. However, as during the Industrial Revolution, the effects of rising productivity, declining distance costs and a new impetus to agglomerations can be observed. The 'steam-oriented' technology and the effects of electricity had an immense impact on physical and economic space, whereas the influence of ICT on economic space (or networks) and physical space is more subtle. The relations are often not direct but 'intermediated' by other variables. For instance, technology leads to new products and to new methods of transportation. The lower distance costs enable concentration in space and agglomeration economies, which in turn can increase productivity and spin-offs. The concentration can also lead to the rise of unexpected forms of creativity and new forms of organisation. The new networks of economic relations can develop without direct spatial contacts. New global structures emerge and agglomerations seem to receive new and stronger impulses, not weaker as was predicted by various authors in the 1980s. The cities develop into networked communities in polycentric spatial patterns.

The relations between technological, economic and spatial processes were often not direct but indirect. The different periods of development of technology, economy and space show different patterns of ordering. Technology and economy evolve more clearly in step with each other, in contrast to spatial patterns that show continuous urbanisation processes, although in periods of considerable changes in technology the growth could stall, more in particular in cities depending on 'old' technologies. More recently, the global spatial structure has undergone very important new shifts in the locations of major corporations, with the rise of new urban growth patterns in Latin America and Asia. This kind of co-evolution can only be analysed and 'explained' over longer time-horizons.

Bibliography

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

Anderson, P., K.J. Arrow and D. Pines (eds) (1988), *The Economy as an Evolving Complex System*, Reading, MA: Addison-Wesley.

Andersson, A.E. (1981), 'Structural change and technological development', *Regional Science and Urban Economics*, **11**, 351–61.

Antonelli, C. (2003), *The Economics of Innovation, New Technologies and Structural Change*, London: Routledge.

Antonelli, C. (2008), *Localised Technical Change. Towards the Economics of Complexity*, London: Routledge.

Arthur, W.B. (1994), *Increasing Returns and Path Dependence in the Economy*, Ann Arbor: University of Michigan Press.

Arthur, W.B. (1999), 'Complexity and the economy', *Science*, **284**, 107–10.

Bartley, K.F. (2006), 'Technology and the convergence of U.S. urban migration patterns: 1970–2000', *Growth and Change*, **37**, 82–106.

Boschma, R.A. (1994), *Looking through the Window of Locational Opportunities. A Long-Term Spatial Analysis of Techno-Industrial Upheavals in Great Britain and Belgium*, Rotterdam: Tinbergen Institute.

Boschma, R.A. (2004), 'The competitiveness of regions from an evolutionary perspective', *Regional Studies*, **38**, 1001–14.

Boschma, R.A. (2005), 'The role of proximity and performance', *Regional Studies*, **39**, 41–5.

Boschma, R.A. and K. Frenken (2006), 'Why is economic geography not an evolutionary science? Towards an evolutionary economic geography', *Journal of Economic Geography,* **6**, 273–302.

Boschma, R.A. and R.C. Kloosterman (eds) (2005), *Learning from Clusters: A Critical Assessment from an Economic-Geographical Perspective*, Dordrecht: Springer.

Boschma, R.A. and J.G. Lambooy (1999), 'Evolutionary economics and economic geography', *Journal of Evolutionary Economics*, **9**, 411–29.

Boschma, R.A. and J.G. Lambooy (2000), 'The prospects of an adjustment policy based on collective learning in old regions', *GeoJournal*, **49**, 391–9.

Cairncross, F. (2001), *The Death of Distance 2.0. How the Communications Revolution Will Change Our Lives*, London: Texere Publishers.

Caldwell, B. (2004), *Hayek's Challenge: An Intellectual Biography of F.A. Hayek*, Chicago: University of Chicago Press.

Castellacci, F. (2007), 'Technological regimes and sectoral differences in productivity growth', *Industrial and Corporate Change*, **16**, 1105–45.

CBS (2003), *De digitale economie 2003* [The Digital Economy 2003], Voorburg: Centraal Bureau voor de Statistiek.

Chandler, A.D. (1962), *Strategy and Structure: Chapters in the History of Industrial Corporations*, Cambridge, MA: MIT Press.

Chandler, A.D. and B. Mazlish (2005), *Leviathans: Multinational Corporations and the New Global History*, Cambridge: Cambridge University Press.

Cohen, H.F. (1994), *The Scientific Revolution*, Chicago: University of Chicago Press.

Cohen, H.F. (2007), *De herschepping van de wereld* [How Modern Science Came to the World], Amsterdam: Bert Bakker.

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

Doi, J. and K. Mino (2005), 'Technological spillovers and patterns of growth with sector-specific R&D', *Journal of Macroeconomics*, **27**, 557–78.

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

Dosi, G., C. Freeman, R. Nelson, R. Silverberg and L. Soete (eds) (1988), *Technical Change and Economic Theory*, London: Pinter.

Florida, R. (2002), *The Rise of the Creative Class*, New York: Basic Books.

Frenken, K. (2006), 'Technological innovation and complexity theory', *Economic Innovation and New Technology*, **15**, 137–55.

Gay, B. and B. Dousset (2005), 'Innovation and network structural dynamics: study of the alliance network of a major sector of biotechnology industry', *Research Policy*, **34**, 1457–75.

Geels, F.W. (2002), 'Technological transitions as evolutionary reconfiguration processes: a multilevel perspective and a case-study', *Research Policy*, **31**, 1257–74.
Gill, I. and H. Kharas (2009), 'Gravity and friction in growing East Asia', *Oxford Review of Economic Policy*, **25**, 190–204.
Giaoutzi, M. (1989), 'Telecommunications, infrastructure and regional development', in K. Peschel (ed.), *Infrastructure and the Space-Economy*, Berlin: Springer Verlag, pp. 116–30.
Glaeser E.L. and J.D. Gottlieb (2006), 'Urban resurgence and the consumer city', *Harvard Institute of Economic Research, Discussion Paper 2109*.
Gottmann, J. (1961), *Megalopolis: The Rise of the North American Eastern Seaboard*, Cambridge, MA: MIT Press.
Graham, S. and S. Marvin (2001), *Splintering Urbanism. Networked Infrastructures, Technological Mobilities and the Urban Condition*, London: Routledge.
Hayek, F.A. (1945), 'The use of knowledge in society', *American Economic Review*, **35**, 519–30.
Hellyer, M. (ed.) (2003), *The Scientific Revolution*, Oxford: Blackwell Publishing.
Helpman, E. (2004), *The Mystery of Economic Growth*, Cambridge, MA: The Belknap Press of Harvard University.
Huber, F. (2007), *Social Networks and Knowledge Spillovers*, Frankfurt: Peter Lang Verlag.
Jacobs, J. (1969), *The Economy of Cities*, New York: Vintage Books.
Jacobs, J. (1984), *Cities and the Wealth of Nations*, New York: Vintage Books.
Kim, S. (2005), 'Industrialization and urbanization: did the steam engine contribute to the growth of cities in the US?', *Explorations in Economic History*, **42**, 586–98.
Kleinknecht, A. (1987), *Innovation Patterns in Crisis and Prosperity*, London: Macmillan.
Kondratieff, N.D. (1926), 'Die langen Wellen der Konjunktur' [The long waves of economy], *Archiv fur Sozialwissenschaft und Sozialpolitik*, December, 573–609.
Krugman, P. (1995), *Development, Geography and Economic Theory*, Cambridge, MA: MIT Press.
Lambooy, J.G. (1984), 'The regional ecology of technological change', in J.G. Lambooy (ed.), *New Spatial Dynamics and Economic Crisis*, Tampere, Finland: Finnpublishers, pp. 63–77.
Lambooy, J.G. (1987), 'Informatisation and internationalization', *Revue d'Economie Regionale et Urbaine*, **6**, 719–31.
Lambooy, J.G. (1989), 'Meso-economics and organizational ecology', in K. Peschel (ed.), *Infrastructure and the Space-economy*, Berlin: Springer Verlag, pp. 254–73.
Lambooy, J.G. (2002), 'Space for complexity', Farewell Speech, Utrecht University, available at: http://econ.geo.uu.nl/lambooy/lambooy.html.
Lambooy, J.G. (2006), 'Knowledge dissemination and innovation in urban regions', in F.J. Carillo (ed.), *Knowledge Cities*, Amsterdam: Elsevier, pp. 223–33.
Lambooy, J.G. and R.A. Boschma, (2001), 'Evolutionary economics and regional policy', *Annals of Regional Science*, **35**, 113–33.
Lambooy, J.G. and F.G. van Oort (2005), 'Agglomerations in equilibrium?', in S. Brakman and H. Garretsen (eds), *Location and Competition*, London: Routledge, pp. 60–81.
Leidig, G. (2002), 'Raumplanung und Naturwissenschaften, dargestellt am Beispiel von Emergenz- und Komplexitätstheorie' [Spatial planning and science, using emergency and complexity theory], *Gesellschaft für Regionalforschung Seminarberichte*, **45**, 153–76.
Lewis, W.W. (2004), *The Power of Productivity*, Chicago: University of Chicago Press.
Lipsey, R.G., K.I. Carlaw and C.T. Bekar (2005), *Economic Transformations: General Purpose Technologies and Long Term Economic Growth*, Oxford: Oxford University Press.
Maddison, A. (1991), *Dynamic Forces in Capitalist Development: a Long-Run Perspective*, Oxford: Oxford University Press.
Malerba, F. (2006), 'Innovation and the evolution of industries', *Journal of Evolutionary Economics*, **16**, 3–23.
Mandel, E. (1975), *Late Capitalism*, London: New Left Books.
Martin, R. (1999), 'The new geographical turn in economics: some critical reflections', *Cambridge Journal of Economics*, **23**, 65–91.
Martin, R. (2006), 'Path dependence and the economic landscape', in C. Berndt and J. Glückler (eds), *Denkanstöße zu einer anderen Geographie der Őkonomie* [Reflections on Heterodox Economic Geography], Bielefeld: Bielefeld Verlag.
Martin, R. and Sunley (2006), 'Path dependence and regional development', *Journal of Economic Geography*, **6**, 395–437.
Mensch, G. (1975), *Das Technologische Patt* [Technological Stalemate], Frankfurt: Umschau Verlag.
Metcalfe, J.S. (1994), 'Evolutionary economics and technology policy', *The Economic Journal*, **104**, 931–44.
Metcalfe, J.S., J. Foster and R. Ramlogan, (2006), 'Adaptive economic growth', *Cambridge Journal of Economics*, **30**, 7–32.
Mokyr, J. (2002), *The Gifts of Athena*, Princeton, NJ: Princeton University Press.

Naisbitt, J. (1984), *Megatrends. Ten New Directions Transforming Our Lives*, New York: Bantam Books.
Nelson, R.R. (2005), *Technology, Institutions, and Economic Growth*, Cambridge, MA: Harvard University Press.
Nelson, R.R. and S.G. Winter (1977), 'In search of a useful theory of innovation', *Research Policy*, **6**, 36–76.
North, D.C. (1990), *Institutions, Institutional Change, and Economic Performance*, Cambridge: Cambridge University Press.
North, D.C. (1994), 'Economic performance through time', *American Economic Review*, **84**, 359–68.
Pasinetti, L.L. (1981), *Structural Change and Economic Growth*, Cambridge: Cambridge University Press.
Peneder, M. (2002), *Entrepreneurial Competition and Industrial Location*, Cheltenham, UK and Northampton, MA, USA: Edward Elgar.
Perroux, F. (1950), 'Economic space: theory and applications', *The Quarterly Journal of Economics*, **64**, 89–104.
Peschel, K. (1989), *Infrastructure and the Space-economy*, Berlin: Springer Verlag.
Romer, P.M. (1986), 'Increasing returns and long-run economic growth', *Journal of Political Economy*, **94**, 1002–37.
Rosenberg, N. and M. Trajtenberg (2004), 'A general purpose technology at work: the Corliss steam engine in the late nineteenth century in the United States', *Journal of Economic History*, **64**, 61–99.
Rostow, W.W. (1978), *The World Economy*, London: Macmillan.
Sassen, S. (1991), *The Global City*, Princeton, NJ: Princeton University Press.
Saxenian, A. (2006), *The New Argonauts*, Cambridge, MA: Harvard University Press.
Schumpeter, J.A. (1912 [1934]), *Theory of Economic Development*, Cambridge, MA: Harvard University Press.
Schumpeter, J.A. (1939), *Business Cycles*, New York: McGraw-Hill.
Schumpeter, J.A. (1943), *Capitalism, Socialism and Democracy*, London: Allen & Unwin.
Scott, A.J. (2006), 'Creative cities: conceptual issues and policy questions', *Journal of Urban Affairs*, **28**, 1–17.
Solow, R.M. (1957), 'Technical change and the aggregate production function', *Review of Economics and Statistics*, **39**, 214–31.
Solow, R.M. (2003), 'Foreword', in G. McMahon and L. Squire (eds), *Explaining Growth: A Global Research Project*, Basingstoke: Macmillan.
Spufford, P. (2006), 'From Antwerp and Amsterdam to London: the decline of financial centres in Europe', *De Economist*, **154**, 143–75.
Storper, M. and A.J. Venables (2002), 'Buzz: the economic force of the city', paper presented at DRUID Conference, Copenhagen.
Sutton, J. (1991), *Sunk Costs and Market Structure: Price Competition and the Evolution of Concentration*, Cambridge, MA: MIT Press.
Toffler, A. (1980), *The Third Wave*, New York: Bantam Books.
van Duijn, J.J. (1979), *De lange golf in de economie*, Assen: Van Gorcum.
van Duijn, J.J. (1983), *The Long Wave in Economic Life*, London: Macmillan.
van Duijn, J.J., and J.G. Lambooy (1982), *Technological Innovation and Regional Economic Growth: A Meso-Economic Analysis*, Amsterdam: Faculty of Economics Research Memorandum, No. 8207.
van Gelderen, S. (under the pseudonym of J. Fedder) (1913), *Springvloed. De Nieuwe Tijd* [Spring Flood. The New Time], Amsterdam: Fortuyn.
van der Knaap, G.A. (1978), *A Spatial Analysis of the Evolution of an Urban System*, Utrecht: Doctoral dissertation, University of Rotterdam.
van Oort, F.G. (2003), *Innovation and Growth*, London: Ashgate.
Wenting, R. and R.A. Boschma (2006), 'The spatial evolution of the British automobile industry', in A. Reggiani and P. Nijkamp (eds), *Spatial Evolution and Modelling*, Cheltenham, UK and Northampton, MA, USA: Edward Elgar, pp. 267–93.
Weterings, A. (2006), *Do Firms Benefit from Spatial Proximity*?, Utrecht, NGS-series No. 336.

23 The information economy and its spatial evolution in English cities

James Simmie

1. Introduction

This chapter discusses and analyses some of the results of the major paradigm shift to post-industrial capitalism and the service-based knowledge and information economy in English cities. In order to conduct this analysis a case is made for the adoption of an evolutionary economics approach as compared with that of neoclassical economics. This rests on the arguments that a central concern of evolutionary economics is the role of knowledge and information in modern economies and the significance of sources in addition to those provided by markets. It highlights the significance of the uses of knowledge in innovation and consequential movement of firms and industries away from rather than towards some pre-defined equilibrium. It also accepts the reality of bounded rationality.

Technological paradigms and trajectories are among the important paths of change analysed in evolutionary theory. It is argued here that every 50 years or so key innovations lead to major paradigm shifts (Hall and Preston, 1988; Marshall, 1987). During the intervening periods firms and industries proceed along the technological trajectories that are established within the context of those paradigms. The latest of these, dating from around the 1960s, is the major change associated with the development of the knowledge- and information-based economy, the growing importance of services and the relative decline of manufacturing. Despite these major structural changes in modern economies, most of the previous evolutionary analyses of long-term change have tended to focus on manufacturing industries. In contrast the analysis presented here is devoted to explaining the evolution of the English service-based urban system.

These paradigm shifts drive major changes in the division of labour and its distribution within and between spatial economies. Whole new regional economies rise in the course of these changes. Previously one such change was exemplified in the rise of the cotton industry in factories located in the north west of England. Later, another wave of innovations drove the change from fruit farming to computer manufacture in Silicon Valley. More recently, in England, the rise of the information economy and its associated knowledge-intensive business services (KIBS) has been highly concentrated in London and the greater south east of the country. Following these types of change, it is argued here that a critical task for evolutionary economic geography is to describe, typify and explain the spatial evolution of the knowledge and information service-based city economies.

There is a considerable consensus on the importance of knowledge and services in the new post-industrial economy (e.g. Bell, 1973; Daniels, 1993; Drucker, 1969; Metcalfe and Miles, 1999; Romer, 1986; and see Anderson et al., 2000). Much of this work is brought up to date in Rooney et al. (2005). It has also made its way into official government policy as exemplified by the DTI (1998): 'Our competitive future: building the knowledge driven economy'. Among the research questions that this literature has given rise to, with respect to this chapter, is what is the role of and what are the mechanisms for the transfer of knowledge within and between spatial economies. Here it is argued that a contribution to this task can be made by analysing the roles of specific sectors such as KIBS and of networks. Research on the changes taking place in KIBS in the 56 largest cities in England is reported to address the first of these tasks. Research on the existence and development of regular business-to-business linkages that form relatively persistent networks in the City of London and the greater south east is reported with respect to the second. Both show a distinctive spatial distribution of the mechanisms for knowledge transfers.

The research reported leads to the conclusion that the spatial development of the knowledge and information economy is likely to be cumulative and circular (Myrdal, 1957). Thus the regions like the greater south east in England that start with most of the assets required in the new paradigm are likely to end up with more of them at a later date. This process then becomes circular as the greater concentration of desirable assets at this later time forms the basis for further cumulative gains in regions like the greater south east in the longer run. As a result spatial economies are more likely to diverge in the information economy than to converge towards some equilibrium path as predicted by neoclassical theory.

In addition to this introduction and a set of conclusions the chapter is divided into seven substantive parts. These deal respectively with the reasons for adopting an evolutionary approach to economics, basic concepts in evolutionary economic theory, knowledge, information, innovation and equilibrium, paradigm change, the information economy and the spatial division of labour, knowledge spillovers and key sectors, the role of KIBS, and knowledge spillovers, networks and spatial nodes.

2. Why adopt an evolutionary approach to economics?

Evolutionary economics as proposed by Nelson and Winter (1982) and developed by Dosi et al. (1988) offers an alternative and more 'realistic' theoretical approach to the understanding of economic change in general and the roles of technological innovation in particular than that proposed by neoclassical economic theory. It deals specifically with some of the main inadequacies of neoclassical theory. Principal among these are the over-importance attached to markets as sources of information, the unrealistic assumptions concerning the existence of equilibrium conditions and the neglect of the bounded rationality problem.

Briefly, the main differences between evolutionary and neoclassical economics on these three points are first that it is argued in the former that prices and markets are by no means the only social mechanisms that actively transmit economic information (Nelson and Winter, 1982, p. 403). This is particularly important at critical eras of change in technological paradigms. Here markets do not usually instigate the innovations on which they are based and are therefore seldom the prime movers (Dosi, 1982).

Second, one of the main reasons for the inadequacy of neoclassical market-based assumptions is the importance attached to the notion that firms and economies are dynamically inclined to move towards equilibrium conditions. In contrast evolutionary theory suggests that the explanation for economic change should focus on the conditions where temporary convergence towards equilibrium is disturbed by endogenously determined innovative firm behaviour (Nelson and Winter, 1982). Far from being a marginal activity, such deviant firm behaviour should be regarded as the fundamental driving force underlying economic development. In such instances divergence from industrial trajectories could be seen as more 'desirable' than path dependence and convergence.

The combination of the importance attached to markets and the notion that there is some objectively 'best' set of equilibrium conditions to which perfectly competitive markets drive individual firms also leads to inadequate policy assumptions. The most important of these is the idea of 'market failure'. This is based on the assumption that there is a set of optimum equilibrium conditions in abstract market systems that can be divined for firms and industries. As a result of following this assumption, much government economic policy-making in the UK has focused on devising institutional structures to deal with so-called 'market failures' while at the same time assuming that the rest of the national economy is functioning more or less in line with market demands. Such assumptions do not carry policy analysis very far because market failure is ubiquitous (Nelson and Winter, 1982, p. 394).

Evolutionary theory suggests that instead of concentrating on market failure, economic policy analysis should be much more concerned with the problems associated with adjusting to continual change. This would include adapting not only to the 'gales of creative destruction' caused by the periodic major shifts in technological paradigms but also to the changing trajectories encompassed within these long waves of economic change. In practical terms this would shift the focus away from market failure towards identifying market opportunities within these changing conditions.

A third important difference between neoclassical and evolutionary economic theory is the significance attached to bounded rationality (Simon, 1955). Where neoclassical economics deals with technical change at all it has repressed the bounded rationality problem (Nelson and Winter, 1982, p. 40). It is clearly incorrect to assume that, particularly in the case of technical change, there is a complete set of information available somewhere to any firm willing to devote the effort to tracking it down. It is more realistic to assume that all firms have to operate as best they can with partial knowledge. Technical change is then driven by what firms can know at any particular time. There is therefore no reason to suppose that the resulting innovations and trajectory of change will necessarily be optimum. We should generally expect satisficing rather than optimising behaviours from firms in the face of the impossibility of knowing everything about any given set of supply and demand conditions.

As a result of this discussion it is argued here that there are at least three important reasons for adopting an evolutionary perspective rather than the traditional neoclassical stance. These are that the former offers a better approach to understanding both gales of creative destruction and incremental, particularly technological, change, better than market analysis. Second it raises the significance of understanding the causes of disequilibrium rather than abstract equilibrium conditions. Finally it starts with the assumption of bounded rationality. This is much more akin to the circumstances that most firms find themselves in as compared with assumptions of optimality.

3. Basic concepts in evolutionary microeconomic theory

Nelson and Winter (1982) developed three basic concepts of firm behaviour as the basis of their evolutionary theory of economic change. These were routine, search and selection. They start from the assumption that organisations have a tendency to prefer to do the same things in a constant environment than to change. They tend to be much better at self-maintenance and changing in the direction of more of the same than they are at any other kind of change.

This pattern of behaviour is built up not on the basis of a grand rule of optimising behaviour founded on utility maximisation but on the basis of satisficing (Simon, 1955) regular routines. These include well-specified technical routines for producing things through procedures for hiring and firing, ordering new inventory, or increasing production in items in high demand to policies for investment, research and development, advertising and business strategies (Nelson and Winter, 1982, p. 14).

Routines are organisational – the competences, capabilities and skills of firms as a whole. They are built around the division of labour within firms and hence the division of skills between workers in a firm. The skills basis of a firm consists in large part of the experience and tacit knowledge of the workforce. These are hard to codify (Boschma and Frenken, 2006, p. 5). The importance of these kinds of knowledge, how they are acquired and transferred, has become a central concern of both evolutionary economics and geography. Furthermore, the importance of the division of labour within firms and industries in the acquisition and transfer of knowledge should also be noted at this stage and is discussed below.

Once a set of routines has been established, the progress of firms and industries employing those routines tends to become path dependent. In these circumstances, other things being equal, behaviours will continue along well-established trajectories. In the worst case scenario this can lead to lock-in and continuation along paths that are increasingly unsuited to market conditions. This is well illustrated by the classic case of the Swiss mechanical watch industry (Maillat et al., 1996).

The second main concept in evolutionary theory that is required to provide chances that routines do not become too path dependent and locked-in to out dated behaviours is search. Searches are the processes by which routines are redirected and changed. Searches involve organisational activities that are associated with the evaluation of current routines and may lead to their modification, drastic change or even complete replacement (Nelson and Winter, 1982, p. 400).

Search introduces a behavioural activity that can be one of the main driving forces for change in firm behaviours. It involves the development of new knowledge both within firms and in relation to their industrial and general environment. Given the presence of bounded rationality, however, there is no guarantee that all searches will discover relevant new knowledge for changing firm routines. They can be undertaken internally by existing functions within firms. But, as is argued below, increases in the division of labour with respect to the knowledge-based economy often lead to search activities being outsourced to specialised knowledge-intensive business services (KIBS).

Selection is the third key concept in evolutionary theory. Along with search, as defined above, it is part of the simultaneous interacting aspects of the evolutionary process. These dynamic processes in which firm behaviour patterns and market outcomes are jointly determined over time are the core concern of evolutionary theory (Nelson and Winter, 1982, p. 19). Search and selection are the joint processes that lead to the ways in which firms and industries evolve, or indeed fail to evolve, over time. They also mean that history matters because as a result of past searches and selections, the condition of an industry in one period provides the basis for its condition in a following period.

The external selection environment of firms and industries means that actual behaviours are codetermined by the interactions between endogenous routines, searches and the external business environment. In particular firms must interact successfully with major exogenous sources of change such as changes in markets and the competitive environment, changes in technology, changes in policy regimes and major economic and cyclical shocks (Martin, in Simmie et al., 2006).

In between gales of creative destruction firms and industries will generally advance along technological trajectories within existing paradigms. These have been defined as 'the pattern of "normal" problem solving (i.e. of progress) on the ground of a technological paradigm' (Dosi, 1982, p. 152). Technological paradigms are discussed below. Suffice it to say at this stage that the establishment of a new paradigm represents major innovative breaks with the past. In between such major 'gales of creative destruction', to use Joseph Schumpeter's (1939) famous phrase, most firms and industries proceed along established technological trajectories. These become cumulative so that advances are related to the position that a firm, industry, region or even country occupies vis-à-vis existing technological frontiers.

The dynamic interactions between firm and industrial routines, their searches and selection environments, have spatial and geographical implications. The fortunes of industrial sectors and the roles of knowledge and innovation are all critical factors in the physical agglomeration of firms and resulting urban and regional growth differences.

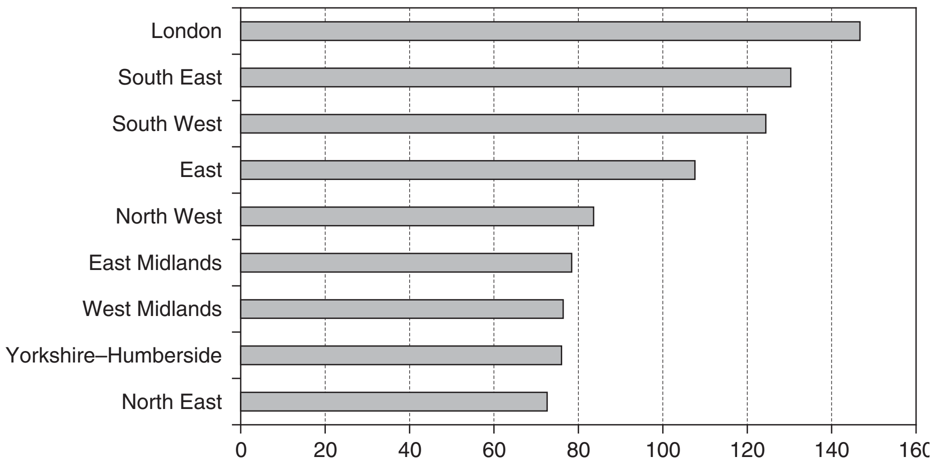
4. Knowledge, information, innovation and equilibrium

Innovation generates disequilibrium. Successful innovation is cumulative (Dosi, 1988; Nelson and Winter, 1982; Ruttan, 1997). R&D investment, for example, is characterised by cumulative complementarities (Bresnahan et al., 2002, Colombo and Delmastro, 2002; Dosi et al., 2003). As a result the firms that can move along existing technological trajectories faster than others can move into oligopolistic positions. Technical change itself becomes a part of the pattern of oligopolistic competition (Dosi, 1982).

In spatial terms, the regions/cities in which most innovation assets are located initially are most likely to accumulate more of them over time. Myrdal (1957) called this the principle of 'circular and cumulative causation'. Kaldor, in making the case for regional policy, argued that differences between regions:

> accrue from the growth of industry itself – the development of skill and know-how; the opportunities for easy communication of ideas and experience; the opportunity of ever-increasing differentiation and of specialisation in human activities. (Kaldor, 1970, p. 340)

Following this line of reasoning it may also be argued that cities (or regions) that have 'first mover advantage' in terms of innovation and technology are likely to attract educated labour and capital from elsewhere, thereby producing a cumulative, self-reinforcing process of research and development leadership. In this sense, therefore, regions/cities may well be in direct competition for better-educated workers and risk capital (Martin, in Simmie et al., 2006). Those spatial economies that are able to attract high quality labour and risk capital and that, as a consequence, are able to proceed rapidly along contemporary technological trajectories are likely to diverge from, rather than converge with, those spatial economies that cannot do this.



Note: UK = 100; year 2000.

Source: Martin (2004).

Figure 23.1 *Index of knowledge-based business*

Figure 23.1 illustrates the lead that southern regions in England had built by 2000 over those in the north with respect to knowledge-based businesses. These are identified according to the proportions of managerial, professional and technical workers they employ. This lead has been established cumulatively over the years principally since the 1960s. Despite government aspirations to the contrary it is difficult to see how, in the kind of circumstances outlined above, the economies of those regions in the north can be made to converge with those in the south given the overall significance of the contemporary knowledge-based economy.

5. Paradigm change

In addition to the evolutionary argument that firms, sectors and economies 'normally' evolve along established trajectories, are path dependent and marked by incremental innovation, they are also subject to gales of creative destruction. These are driven by breakthrough innovations in particular economic sectors that then spread through other parts of the economy. Breakthrough innovations are also situated in a combination of economic, institutional and social factors that contribute to the establishment of new technological paradigms (Dosi, 1982, p. 153). Dosi defines a technological paradigm 'as a model and a pattern of solution of selected technological problems, based on selected principles derived from natural sciences and on selected material technologies' (Dosi, 1982, p. 152).

Each new technological paradigm has drastic organisational and geographical implications. In the UK, the first modern industrial wave, based on key innovations in weaving, iron manufacture and transportation, drove the development of the factory system, a laissez-faire economic regime and the new growth of towns around coalfields and ports. The second wave, based on key innovations in steel and steamships, was marked by larger factories, capital concentration and the development of the joint stock company. Geographically, in the UK, it also led to the growth of large towns particularly around coalfields. The third wave, based on innovations in electricity and automobiles, produced the giant factories of 'Fordism', cartels and the growth in the significance of finance capitalism. Geographical results included the forming of large conurbations and the embryonic development of the international economy. The fourth wave was driven by innovations in computing and communications and characterised by a mixture of large 'Fordist' and small subcontracting factories combined with the emergence of large multinational companies. Geographically it drove suburbanisation, deurbanisation and the establishment of new industrial regions such as Silicon Valley. It was also associated with a growing international division of labour (Hall and Preston, 1988).

As part of this fourth major paradigm shift it has been argued that the advanced economies are in the throes of an accelerated transition from 'industrial' to 'post-industrial' capitalism. This acceleration was recognised by scholars such as Daniel Bell (1973) in his seminal work on 'The Coming of Post-industrial Society', on the cover of the Penguin edition of which is scrawled 'knowledge rules O.K.'. Bell himself later changed the title to the 'information society' for which concept he is generally regarded as the originator.

The information society is characterised by:

- The expansion and dominance of service sectors in advanced economies.
- The new technological paradigm based on the production, processing, communication and consumption of information.
- New information and communication enabling technologies, a growing wave of globalisation and the consequential significance of the international economy.

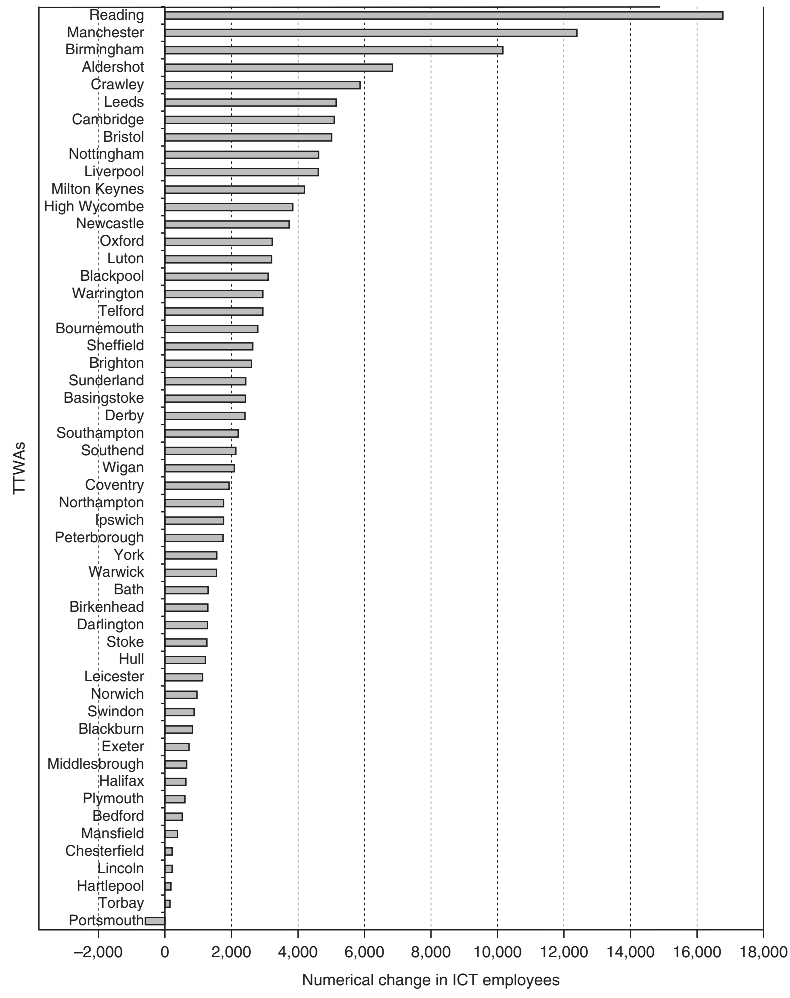
An important task for evolutionary economic geography is therefore to describe, typify and explain the spatial emergence and evolution of the knowledge-driven information society. In principle, evolutionary models attempt to make such explanations in terms of the geographical patterns of industries and the ways in which they acquire and exploit knowledge in conditions of uncertainty and bounded rationality. These analyses at the level of firms and industries can be built into an explanation of the spatial evolution of an economic system as a whole in terms of a multi-sector framework of structural change and networks located in space (Boschma and Frenken, 2006; Dosi and Soete, 1988; Hall and Preston, 1988; Hohenberg and Lees, 1995; Pasinetti, 1993).

6. The information economy and the spatial division of labour

Information is a key evolutionary dynamic:

> Information about new scientific developments, information regarding the success or failure of R&D projects to guide the next round of R&D decisions, information regarding the characteristics of new products to guide potential purchasers, information regarding costs of production and purchases to guide producers, information about profits to guide investors. (Nelson and Winter 1982, p. 29)

All these are critical inputs to the evolutionary processes of routines, search and selection.



Note: Definition of computer and related ICT services – 1992 SIC category 72.

Figure 23.2 Absolute change in ICT employees 1991–2002 (excluding London)

In England, information systems and the kinds of service sector whose main functions are the enabling and exchange of information have developed and evolved at different rates in different locations. Recent research (Simmie et al., 2006), commissioned by the Office of the Deputy Prime Minister (ODPM), has examined the factors underlying the competitiveness of the economies of the 56 largest cities in England with populations exceeding 125,000. The locations and extent of the 'travel to work areas' (TTWAs) of these cities are shown in Appendix 1.

As part of this analysis the rates of change in employment in the ICT sector were examined. Figure 23.2 shows the absolute changes in ICT employment (SIC 72) between 1991 and 2002. The figure excludes London where the growth was 90,806 as compared with the next TTWA, Reading, that recorded a gain of 16,776 employees. It may be seen that while some of the bigger cities such as Manchester, Birmingham, Leeds, Bristol, Nottingham, Liverpool, and Newcastle recorded some of the largest absolute gains, smaller cities in the greater south east recorded high relative rates of growth. Thus Reading, Aldershot, Crawley, Cambridge, Milton Keynes, High Wycombe, Oxford and Luton all posted large increases in ICT employment. This indicates the development of a distinctive spatial division of labour in the sector that plays a key role in providing the infrastructures that facilitate and enable the contemporary exchange of knowledge and information. The dimensions of this spatial division of labour have already been pres-aged in Figure 23.1's analysis of the regional distribution of knowledge-based businesses in 2000.

The pace of change and resulting spatial division of labour in the UK were accelerated in the 1980s by the political economy regime instigated and followed by the Thatcher governments. This regime brought to an end the post-war Keynesian consensus on demand management and counter-cyclical investment. This was replaced by an experiment in Friedmanite monetarism, a refusal to provide sectoral pump priming investment during the recession of the early 1980s, privatisation and the run down of industrial subsidies. This greatly increased the rate of de-industrialisation during that decade. The brunt of this was borne by the older industrial cities of the north and also in London (Martin, 2004).

At the same time an enormous boost was given to the financial services sector by the de-regulation of the City of London, the relaxation of exchange controls and capital movements, and the lowering of taxes that all helped to fuel a boom in London and the greater south east (Martin, 2004). Some of the spatial results of this sectoral restructuring can be seen in Figure 23.3. This shows that the average proportion of service sectors in the English economy had reached 80 per cent of employment by 2000. Of the 17 TTWAs where this proportion was exceeded, only two were located outside the south of England. Conversely of the 39 TTWAs that fell below the English average, 30 were located in the north.

These data also raise questions over the relative scale and importance of high-tech manufacturing in the overall economy. Given that manufacturing as a whole is a decreasing proportion of economic activity in the English economy, and high-tech is only a minority of that decreasing proportion, much is expected from a small part of the total economy. Although the contribution of high-tech manufacturing is crucial in the information economy, it is not a large enough part of the total economy to drive economic growth on its own. Nevertheless, the diffusion of high-tech products, such as ICT, into other sectors of the economy does have significant impacts on productivity and economic growth.

Figure 23.3 also indicates the distinctive spatial division of labour that is emerging in the English information-based service economy. It is also the case that the nature and quality of service employment tend to be different in different cities. In the south east, for example, and particularly London, the new information and telecommunications technologies have enabled the city to take advantage of deregulation and highly profitable electronic trading. In a number of northern cities such as Leeds and Liverpool the new information technology has often facilitated the growth of call centres, the new sweatshops of the service economy. Thus there is a wide variation in the quality

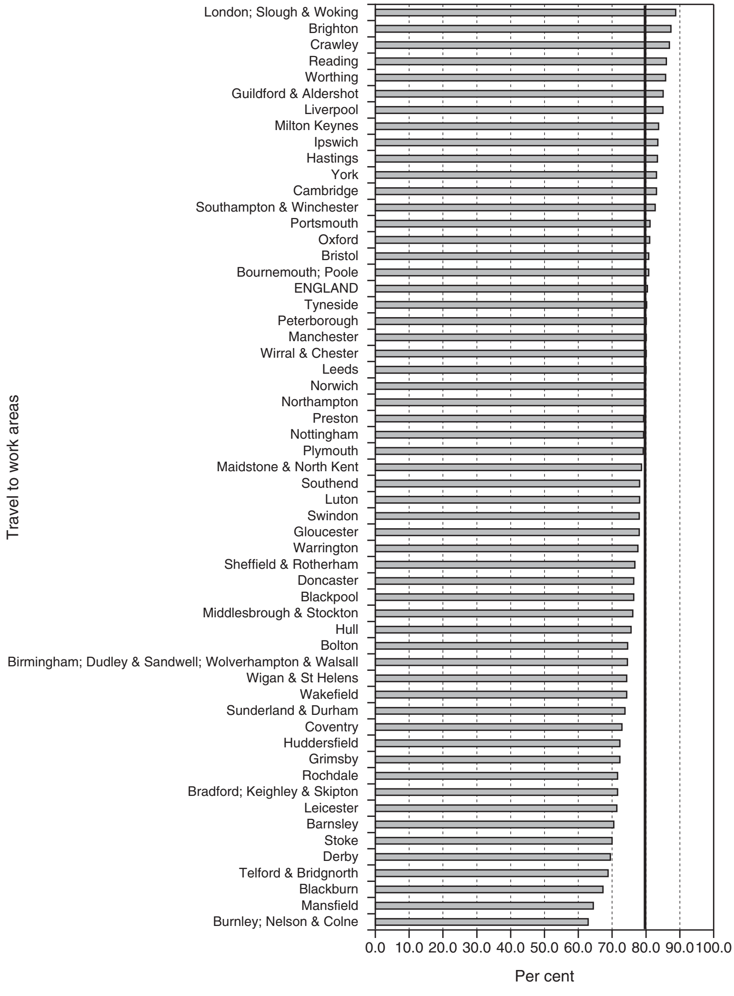


Figure 23.3 Services as a proportion of total employment, 2000

of information-based service employment, both in terms of the overall growth of information-based services and in terms of their quality as exemplified by the growth of well paid KIBS in the greater south east as opposed to the growth of low pay call centres in some northern cities.

7. Knowledge spillovers and key sectors; the role of KIBS

From an evolutionary perspective an important element of the information economy is the spillover of knowledge from one firm or organisation to another. Of particular interest are the ways in which tacit and uncodified knowledge spills over and provides incentives for spatial agglomeration to occur (Jaffe et al., 1993). Tacit knowledge does not travel well and so tends to be spatially concentrated among geographically proximate agents.

The location of certain varieties of firm in particular places and the specific knowledge externalities that are generated as a result tend to have self-reinforcing effects on the selection and nature of firms located in particular spatial economies (Arthur, 1990; Myrdal, 1957). As a result, regional and urban variety in technological trajectories are sustained as knowledge specific to the sectors present spills over primarily among proximate firms (Boschma and Frenken, 2006). Spatial proximity remains significant here despite the growing use of ICT because tacit knowledge is attached to individuals and does not travel easily. Conversely, locations, such as some of those in the north of England, where knowledge-intensive firms are thin on the ground, suffer from a lack of knowledge-rich business environments.

Knowledge spillovers do not take place automatically. They rely on various transfer mechanisms such as imitation, spin-offs, inter-firm collaboration, social and professional networks and labour mobility (Camagni, 1991; Capello, 1999; Boschma and Frenken, 2006; Breschi and Malerba, 2003). There is also a group of sectors whose main function in the division of labour in the information economy is to monitor, develop, combine and transfer knowledge. These are knowledge-intensive business services or KIBS. Their precise definition in terms of NACE codes is shown in Appendix 2.

These sectors have grown rapidly since the mid-1980s. Demand for the kinds of knowledge transfer that they can provide has been driven by the increasingly specialised requirements of businesses in the information economy. In these circumstances KIBS are relatively unique in their ability to appropriate and commercialise certain kinds of knowledge and to transfer them between businesses. They are both a product of the information economy and one of its distinguishing characteristics.

In many of the traditional sectors of knowledge-intensive services, the new ICT technological paradigm brought about huge structural changes. Many of these sectors have been characterised by the combination of previously separate strands of knowledge. Management consultants, for example, have integrated information and communication technology consulting services; advertising agencies offer multimedia services; and technological services such as software firms provide management consulting. Numerous new firms that provide new services that did not exist a few years ago have set themselves up both within and between the sectors and at the interfaces. There are mainly three aspects that provide the links between the heterogeneous KIBS sectors and can be seen as common definitional characteristics:

- Knowledge is not only a key production factor of the firms, it is also the ‘good’ they sell. The firms for the most part provide non-material intangible services. Specialised expert knowledge, research and development ability, and problem solving are the real products of KIBS.
- The provision of these knowledge intensive services requires in-depth interaction between supplier and user and both parties are involved in cumulative learning processes. These learning processes must occur if a transfer of knowledge or a problem solution is to succeed. The utilisation of knowledge intensive services cannot simply be equated with the purchase of standardised external services.
- The third important common aspect of all KIBS sectors is that the activity of consulting, understood as a process of problem solving in which the KIBS adapt their expertise and expert knowledge to the needs of the client, makes up, to different

degrees, the content of the interaction process between KIBS and their customers (Strambach, 2002a, 2002b).

A number of case studies in different KIBS sectors have shown that the critical knowledge transfer functions of KIBS include:

- transferring knowledge in the form of expert technological knowledge and management know-how
- exchanging empirical knowledge and best practice from different branch contexts
- integrating different stocks of knowledge and competencies that exist in innovation systems
- adapting existing knowledge to the specific needs of the client (Bessant and Rush, 2000; Wood, 2002).

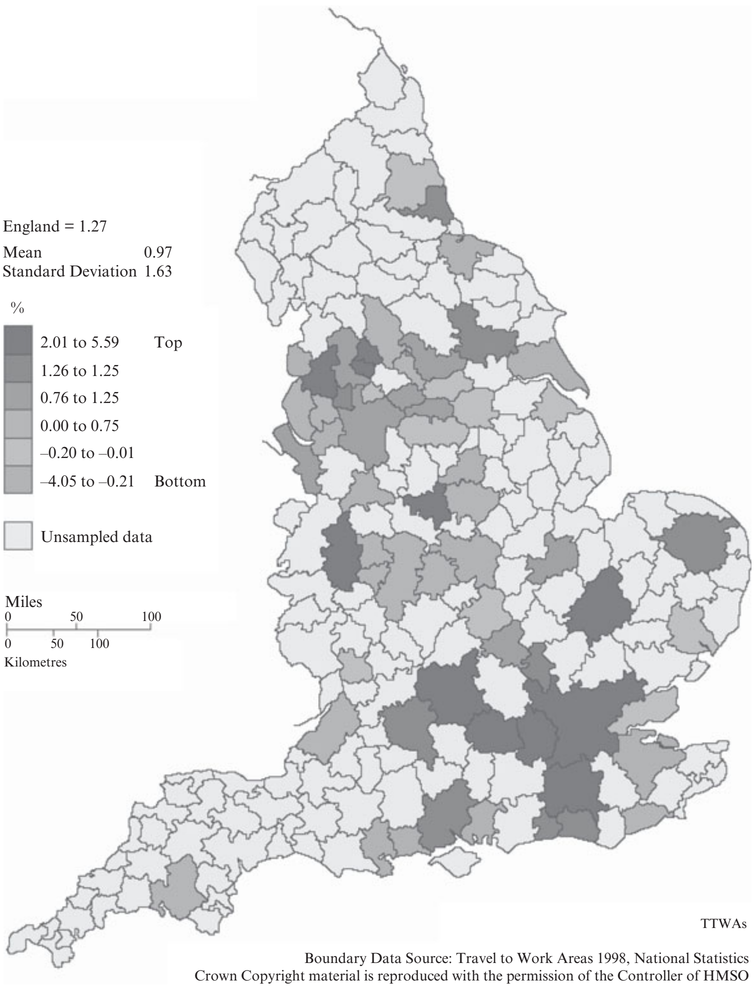
These services are therefore in the forefront of the development of the information economy. The dynamic growth of KIBS is an indication of the presence, in a particular local economy, of a knowledge-rich business environment, a pool of other types of firm that recognise the need for expert external business services and a sophisticated division of labour. Figure 23.4 shows the dynamic changes taking place in KIBS in the economies of English cities during the 1990s. It may be seen that the average growth rate in these sectors in England during the 1990s was 1.27 per cent. Most of the urban economies that exceeded this rate of growth were located in and around London, particularly to the west and south of the metropolis. The local economy of Cambridge was also one of the faster growing areas in the greater south east (GSE). Only four other urban economies outside the GSE experienced similar rates of growth, albeit from a low initial base in 1991. These were Telford and Bridgnorth, Derby, Preston and Burnley.

Figure 23.4 shows clear differences in the spatial dynamics of change in key sectors of the information economy. While London and its western arc running from Surrey through the Thames valley and out to Cambridgeshire were at the forefront of the developing information economy during the 1990s, the growth trajectories in many of England's more peripheral cities and those from the midlands to the north east were not keeping pace with the national shift towards an immaterial knowledge-based economy.

8. Knowledge spillovers, networks and spatial nodes

In the information society, business market links and non-market social networks play crucial roles in knowledge spillovers and transfers. In principle the spatial evolution of networks starts in neutral space in which many locations are candidates to be key nodes in the system (Boschma and Frenken, 2006). But in practice only a few places emerge over time as central nodes with high levels of connectivity to other hubs. Network evolution can be understood as a continually changing landscape of connectivity between emerging and existing nodes, depending on the intensity of their interactions and, to some extent, their geographical distance from each other (Castells, 1996).

Connectivity between nodes tends to be skewed and this contributes to the selection processes between them. This leads to the development of hierarchical relationships between them, with a few locations becoming the best connected primary nodes and others becoming secondary hubs, while most locations develop into less connected



Note: Definition of KIBS – 1992 SIC categories 72, 73, 74.1, 74.2, 74.3, 74.4; figures have been based on sextiles.

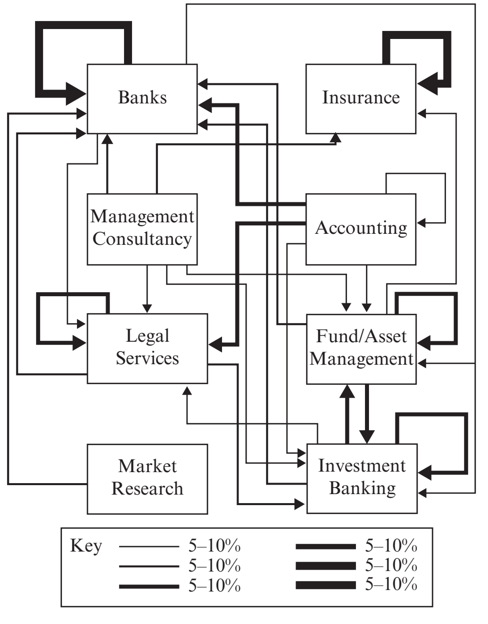
Figure 23.4 Percentage change of employees in knowledge-intensive business services between 1991 and 2001

spokes (Boschma and Frenken, 2006). This places a premium on the continual allocation of resources to network and infrastructural developments such as business and social linkages together with transport and communications. Regions that remain locked in to mature industries and existing physical and social infrastructures are likely to experience structural decline (Boschma and Frenken, 2006).

London and its surrounding areas provide one of the best examples of dynamic and hierarchical network development in the service-based information economy. For many of the firms in the area access to the kinds of concentration of ICT specialists shown in Figure 23.2 is crucial. They are essential to the development of electronic infrastructures, the processing and flows of information that are the lifeblood of the kinds of business service sector concentrated in inner London. Firms in the region also value highly the access to knowledge and knowledge spillovers that follows from locating close to support services, competitors and customers. Banks are key players in this system and were found to value geographic concentration 'to sustain localised formal and informal networks, and social interaction . . . for knowledge accumulation and transfer' (City Corporation of London, 2003, p. 5).

The London labour market is a key mechanism enabling local, national and international knowledge spillovers often through networks. Labour can be recruited through both domestic and international pools. The rewards available in London attract labour from Europe, North America and the Far East. The intellectual quality of much of this labour is a source of knowledge spillovers not just from the UK but from other international locations. Thus new knowledge is brought into the local networks as a result of the attraction of following a career path in London. The sheer size of these kinds of labour market also means that mobility between firms and sectors is encouraged and this is also a major source of local knowledge spillovers (City Corporation of London, 2003, p. 6).

Business and social networks are dense in the City of London. Figure 23.5 illustrates the complexity of network connections within the financial services industry. In this study organisations were asked to say which kinds of firm were most important to them in terms of business interrelationships. The arrows show the direction of the relationship. The thicker the line the more important is the interaction. Briefly, Figure 23.5 shows the significance of banks, legal services and fund management as the hubs of the network.



Source: City Corporation of London (2003).

Figure 23.5 Networking within the London financial services concentration

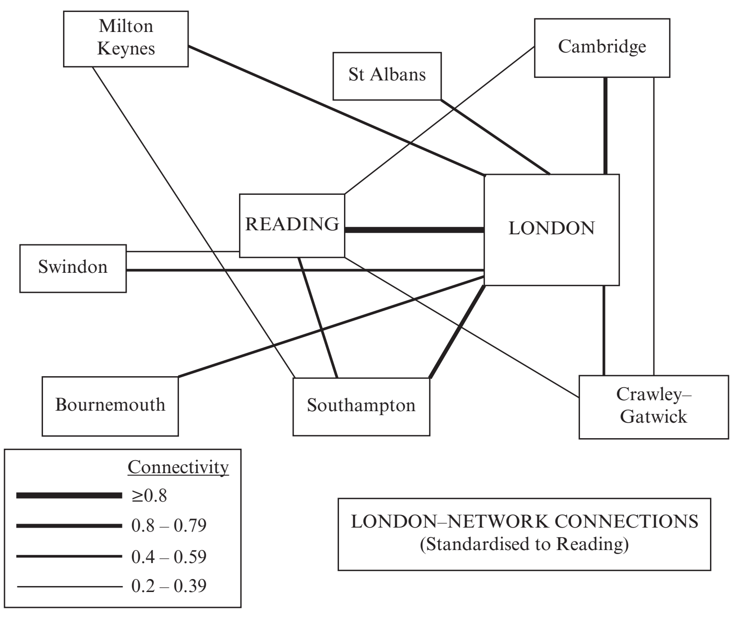
The role of KIBS is also highlighted in terms of management consultancy, accounting and market research.

The importance of KIBS in regional knowledge networks is also shown in Figure 23.6. In this study the locations of the multiple offices of business service firms were analysed with respect to the greater south east (GSE). The strength of these intra-firm networks was then estimated according to the numbers of firms having branches in each of them. These numbers showed a distinctive bias towards the western arc around London. This extended from Cambridge in the north around Swindon in the west and down to Southampton in the south. This geography of business services reflects that shown in Figure 23.4 for the changes taking place in KIBS during the 1990s.

These studies suggest the potential significance of networking between core service sectors and KIBS in the processes of knowledge spillovers in the information economy. They also show the relative importance of a node like the City of London in the hierarchy of nodes at least in its local region the GSE. Identifying the main actors in these nodes and networks is an essential first step in analysing the evolution of the mechanisms that form the bases of knowledge spillovers.

Networked knowledge spillovers are a critical element of interactive learning and innovation systems. While these mechanisms have been studied extensively with respect to innovation in technological product and process innovations, comparatively little is known about innovation in services. Given that services make up a large majority of economic activity in post-industrial information based economies this is a major lacuna.

Nevertheless, in the City Corporation of London (2003) survey of networking in the London financial services concentration, illustrated in Figure 23.5, some familiar results



Source: Institute of Community Studies POLYNET; www.icstudies.ac.uk/html//whatdo_A.asp.

Figure 23.6 Business services networks in the greater south east emerged. The research showed that innovation is critical to the evolution of the sectors. Relatively localised networking between skilled labour, customers and suppliers was an important factor for business transactions and product design. Legal firms rated highly the importance of customers as potential innovators demanding increasingly sophisticated services. In contrast, banks, fund managers and insurance firms rated complementary suppliers, similar to themselves, as driving innovation through competition. It was also found that the larger the firm the more it rated competition as a primary driver of innovation particularly as firms diversified their activities in order to compete for (niche) market shares (City Corporation of London, 2003, p. 7). Networking provided key avenues for knowledge spillovers that enabled firms to learn what their customers required and what their competitors were doing and therefore to be able to adapt their routines and innovate ahead of them.

Competition was found to be one of the main drivers of innovation in the City of London. Banks, and especially investment banks, formed the core sector in the City and Canary Wharf. Their face-to-face meetings within the locality facilitated knowledge spillovers that enabled their searches for innovative product differentiation and new market shares. A City location and networking were particularly important for banks. Both appeared to make important contributions to their abilities to innovate and to adapt their routine behaviours through strategic reorientation and the development of new organisational structures (City Corporation of London, 2003, p. 47).

Another feature of networking in the City was that it enabled the assembly of multidisciplinary teams on a 'pick and mix' basis in order to meet customer needs. Frequently these do not appear as pure legal or pure financing requirements. In these circumstances the ability to bring together complementary expertise to produce novel and bespoke solutions to customers' needs is an important form of incremental innovation in services. While each solution is unlikely to represent a breakthrough innovation, nevertheless, the recombination of previous knowledge and information into new combinations does move the services provided along existing forms of knowledge trajectories.

9. Summary and conclusions

Evolutionary economics offers possibilities for overcoming some of the limitations of the underlying assumptions of neoclassical economics. In particular it addresses the issues of the limitations of the information provided to firms by markets, the role of innovation in driving firms away from equilibrium, and the problems of imperfect knowledge and bounded rationality. In contrast to neoclassical theory, evolutionary theory starts from the proposition that firms normally follow established routines. Over time these can lead to path dependence and lock-in to increasingly outdated practices. The process of searching for adaptations and new routines is the way in which firms can acquire new knowledge and adapt their routines. Selection takes place as searches interact with internal routines and the external environment. Those firms that are best able to adapt their routines to external demands tend to survive more often than those that are less able to deal with change.

The social and spatial divisions of knowledge labour have become much more important with the paradigm shift to post-industrial capitalism or the 'information society'. This is reflected most obviously in the high proportions of services in advanced economies. In these circumstances a main task of evolutionary economic geography is to describe, typify and explain the spatial evolution of the knowledge and information economy.

The new information economy is characterised by an uneven spatial division of labour. ICT is the key technology underlying this economy. The distribution of ICT competences among the workforce in England is concentrated in and around the metropolitan area of London and the greater south east. A part of the explanation for this has to do with changes in the national political economy regime wrought in the 1980s and early 1990s by successive Thatcher governments. These hastened the deindustrialisation of the north and London, while deregulation in financial services stimulated the growth of electronically delivered trading services in the City of London. Part of the result of these changes is the very high level of all services in places like the London TTWA where they are approaching some 90 per cent of the entire local economy.

A characteristic of the evolution of new services is that their quality is unevenly distributed. The spatial division of labour shows highly paid electronic financial traders concentrated in London. In contrast Leeds has become the call-centre capital of England (Leeds City Council, 2007).

Knowledge and information are the key raw materials of the information economy. The agglomeration and concentration of such economies is partly driven by the availability of knowledge spillovers. There are a number of mechanisms by which such spillovers take place. Two of these are the growth of sectors that make their living by transferring knowledge, and the development of various forms of network.

KIBS are an example of the former. They provide commercial, interactive knowledge spillovers between themselves and between knowledge-based and capable businesses in other sectors. They have grown rapidly over the past few decades on the basis of the increasingly sophisticated division and re-division of labour in the information economy. There tends to be a spatially symbiotic relationship between KIBS and other knowledge-based businesses. They co-evolve together and so where KIBS are found it is also to be expected that there will be concentrations of other knowledge-based businesses. In England this is especially the case in the City of London and mainly around the greater south east. These locations provide another example of the uneven spatial division of labour in the information economy.

Networking is also an important mechanism for enabling knowledge spillovers. Dense co-location of complementary knowledge-based sectors and networking is illustrated by the banks and KIBS concentrated within the City of London. These networks also extend beyond the City into Canary Wharf and out into the mainly western arc around the GSE. These strong local networks also form the basis for international knowledge transfers through the interchange of high quality labour facilitated by high capacity ITC infrastructures.

In the City, networking was found to facilitate innovation through knowledge transfers between complementary services and their customers and also as a result of knowledge of competitors. Although innovation in services has proved difficult to conceptualise it can be seen in the recombination of competences and knowledge on a pick and mix basis into new bespoke knowledge for specific customers.

Evolutionary economic theory suggests some relatively uncomfortable realities for firms and sectors. Both are subject to constant competitive selection processes as they seek to adapt their existing routines to proceed along existing technological trajectories.

Those that succeed in doing this tend to prosper while those that fall behind can atrophy and die. The disruptive technologies and processes associated with new technological paradigms confront existing firms with even greater challenges and introduce whole new industries into regional and national economies.

Previous Kondratieff (1935) / Schumpeterian (1939) paradigm shifts have not only had major economic implications but also have had significant spatial results. The economies of whole regions have risen and fallen on the basis of such changes. This highlights the theoretical limitations of the neoclassical assumption that in the long term regional and national economies should be moving towards some equilibrium path. Empirical support for this assumption is either unsatisfactory because of the limitations of the models used to calculate long-term convergence (Henley, 2005), or lacking altogether.

From an evolutionary perspective it is to be expected that as firms and industries move along existing technological trajectories at different speeds the spatial economies in which they are located will also diverge at similar rates. The processes of cumulative and circular causation in which those economies that start with the most relevant economic resources end up with even more of them later in time also contribute to the divergence of spatial economies. Breakthrough innovations leading to major technological paradigm shifts magnify the scale and scope of these selective forces. Thus it is to be expected that the shift to post-industrial capitalism and the development of the information economy will result in the accelerated divergence among spatial economies. This appears to be the case in England where the regions that benefited most from earlier manufacturing growth are now lagging behind those that have participated most and adapted best to the demands of the information economy. Recently this has become a major policy issue, with central government adopting a target to close the gap in growth rates between those of the poorer regions of the north and the faster growth regions in the south. Evolutionary theory and the evidence presented here suggest that without major improvements in the information-based economies of the northern regions this target cannot be attained in any meaningful way.

References

Anderson, B., J. Howells, R. Hull, I. Miles and J. Roberts (2000), *Knowledge and Innovation in the New Service Economy*, Cheltenham, UK and Northampton, MA, USA: Edward Elgar.

Arthur, W.B. (1990), 'Silicon Valley locational clusters: when do increasing returns imply monopoly?', *Mathematical Social Sciences*, **19** (3), 235–51.

Bell, D. (1973), *The Coming of Post-industrial Society: A Venture in Social Forecasting*, New York: Basic Books.

Bessant, J. and H. Rush (2000), 'Innovation agents and technology transfer', in M. Boden and I. Miles (eds), *Services and the Knowledge-Based Economy*, London & New York: pp. 155–69.

Boschma, R.A. and K. Frenken (2006), 'Why is economic geography not an evolutionary science? Towards an evolutionary economic geography', *Journal of Economic Geography*, **6** (3), 273–302.

Breschi, S. and F. Malerba (2003), 'Mobility and social networks: localised knowledge spillovers revisited', CESPRI Working Paper 142, available at: www.cespri.it

Bresnahan, T., E. Brynjolfsson and L.M. Hitt (2002), 'Information technology, workplace organization and the demand for skilled labor: form level evidence', *Quarterly Journal of Economics*, **117**, 339–76.

Camagni, R. (ed.) (1991), *Innovation Networks. Spatial Perspectives*, London & New York: Bellhaven Press.

Capello, R. (1999), 'Spatial transfer of knowledge in high technology milieux: learning versus collective learning processes', *Regional Studies*, **33** (4), 353–65.

Castells, M. (1996), *The Rise of the Network Society*, Oxford: Blackwell.

City Corporation of London (2003), 'Financial services clustering and its significance in London', Report prepared by Manchester Business School and Loughborough Department of Geography, London, City Corporation.

Colombo, M.G. and P. Delmastro (2002), 'The determinants of organizational change and structural inertia: technological and organizational factors', *Journal of Economics and Management Strategy*, **11**, 595–635.

Daniels, P. (1993), *Service Industries in the World Economy*, Oxford: Blackwell.

Department of Trade and Industry (DTI) (1998), *Our Competitive Future: Building the Knowledge Driven Economy*, London: The Stationery Office.

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

Dosi, G. (1988), 'Source, procedure and microeconomic effect of innovation', *Journal of Economic Literature*, **26**, 1120–71.

Dosi, G. and L. Soete (1988), 'Technical change and international trade', in G. Dosi, C. Freeman, R. Nelson, G. Silverberg and L. Soete (eds), *Technical Change and Economic Theory*, London: Pinter, pp. 401–31.

Dosi, G., C. Freeman, R. Nelson, G. Silverberg and L. Soete (eds) (1988), *Technical Change and Economic Theory*, London: Pinter.

Dosi, G., D.A. Levinthal and L. Marengo (2003), 'Bridging contested terrain: linking incentive-based and learning perspectives on organisational evolution', *Industrial and Corporate Change*, **12**, 413–36.

Drucker, P. (1969), *The Age of Discontinuity: Guidelines to our Changing Society*, New York: Harper & Row.

Hall, P. and P. Preston (1988), *The Carrier Wave: New Information Technology and the Geography of Innovation, 1846–2003*, London: Unwin Hyman.

Henley, A. (2005), 'On regional growth convergence in Great Britain', *Regional Studies*, **39** (9), 1245–60.

Hohenberg, P.M. and L.H. Lees (1995), *The Making of Urban Europe*, Cambridge, MA: Harvard University Press.

Jaffe, A.B., M. Trajtenberg and R. Henderson (1993), 'Geographic localization of knowledge spillovers as evidenced by patent citations', *Quarterly Journal Journal of Economics*, **108** (3), 577–98.

Kaldor, N. (1970), 'The case for regional policy', *Scottish Journal of Political Economy*, **17** (3), 337–48.

Kondratieff, N.D (1935), 'The long waves in economic life', *Review of Economic Statistics*, **17** (6), 105–15.

Leeds City Council (2007), *Directory of Call Centres in Leeds and Yorkshire*, Leeds: LCC.

Maillat, D., G. Léchot, B. Lecoq and M. Pfister (1996), *Comparative Analysis of the Structural Development of Milieux: The Example of the Watch Industry in the Swiss and French Jura Arc*, Neuchâtel, Switzerland: IRER-University of Neuchâtel.

Marshall, M. (1987), *Long Waves of Regional Development*, London: Macmillan.

Martin, R.L. (2004), 'The contemporary debate over the North–South divide: images and realities of regional inequality in late-twentieth century Britain', in A.R.H. Baker and M.D. Billinge (eds), *Geographies of England: The North–South Divide, Imagined and Material*, Cambridge: Cambridge University Press, pp. 15–43.

Metcalfe, J.S. and I. Miles (eds) (1999), *Innovative Systems in the Services Economy*, London: Kluwer.

Myrdal, G. (1957), *Economic Theory and Underdeveloped Regions*, London: Gerald Duckworth.

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

Pasinetti, L.L. (1993), *Structural Economic Dynamics*, Cambridge: Cambridge University Press.

Romer, P.M. (1986), 'Increasing returns and long-run growth', *Journal of Political Economy*, **94** (5), 1002–37.

Rooney, D., G. Hearn and A. Ninan (2005), *Handbook on the Knowledge Economy*, Cheltenham, UK and Northampton, MA, USA: Edward Elgar.

Ruttan, V.W. (1997), 'Induced innovation, evolutionary theory and path dependence: sources of technical change', *Economic Journal*, **107**, 1520–29.

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

Simmie, J.M., J. Carpenter, A. Chadwick, R.L. Martin and P. Wood (2006), *The Competitive Economic Performance of English Cities*, London: ODPM.

Simon, H.A. (1955), 'A behavioural model of rational choice', *Quarterly Journal of Economics*, **6**, 99–118.

Strambach, S. (2002a), 'Organisational innovations in different systems of innovation – the case of Germany and UK', in L. Schätzl and J. Revilla Diez (eds), *Technological Change and Regional Development in Europe*, Heidelberg & New York: Springer Physica Verlag, pp. 250–72.

Strambach, S. (2002b), 'Germany – knowledge-intensive services in a core industrial economy', in P. Wood (ed.), *Consultancy and Innovation: The Business Service Revolution in Europe*, London: Routledge, pp. 124–51.

Wood, P. (2002), 'Knowledge intensive services and urban innovativeness', *Urban Studies*, **39**, (5–6), 993–1002.

Appendix 1

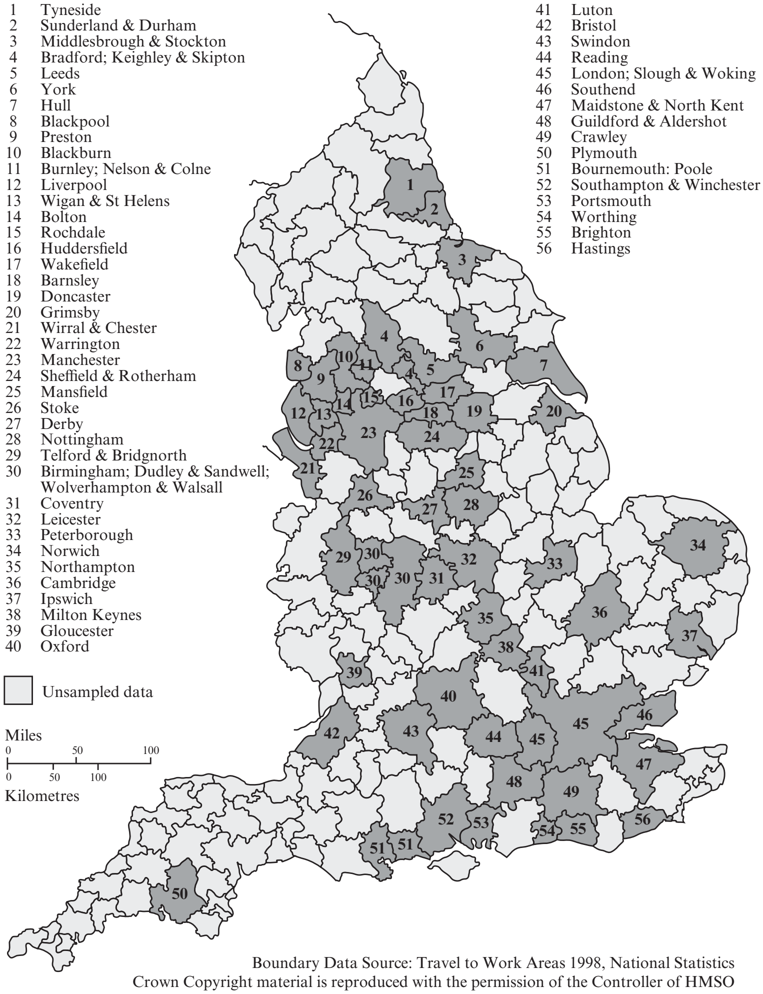


Figure 23A.1 Index map of the 56 TTWAs used in this study

Appendix 2

Table 23A.1 Narrow definition of KIBS

NACE	Branch	Branch aggregation
72	**Computer and related services**	**Computer services**
72.1	Hardware consultancy	
72.2	Software consultancy and supply	
72.3	Data processing	
72.4	Database processing	
72.5	Maintenance, repairing of office machines	
72.6	Other services related with data processing	
73	**Research and development**	**Research & development**
73.1	Research and development in the natural sciences	
73.2	Research and development in social, economic sciences	
74	**Other business activities**	**Economic services**
74.1	Legal services, tax consultancy, management consultancy, market research . . .	
74.2	Architecture and engineering offices	**Technical services**
74.3	Technical, physical, chemical analysis	
74.4	Advertising	**Advertising**

24 An evolutionary model of firms' location with technological externalities*

Giulio Bottazzi and Pietro Dindo

1. Introduction

The present contribution intends to pursue the analysis of the effects of the evolutionary metaphor (Dosi, 1988; Freeman, 1986; Nelson and Winter, 1982) when applied inside the domain of economic geography. In principle, the validity of the assumption of evolutionary economics is all but obvious and the question of whether economic interactions can be effectively thought of as an evolutionary process still open. With the clear risk of oversimplifying the matter, we could say that the notion of evolution immediately entails three consequences for the economic dynamics. First, it should move from simpler to more complex structures. Second, it should progressively eliminate less efficient structures and promote the development of more efficient ones, irrespectively of the fact that this process of elimination and promotion might take place through a mechanism of adaptation by some of the economic actors or through an 'adoptation' by some of the markets and institutions (Alchian, 1950). Third, the progressive change or renewal of the different actors and rules should proceed in a jointly integrated way.

Obviously, the central question is not whether the characteristics described above can be considered to be present in economic systems, because they certainly are. The question is whether the evolutionary accounting of their effects and causes allows for a deeper understanding and a more reliable modeling of economic interactions. In the end, one is interested to know if this accounting could help in the development of more effective policies. In principle, however, the ideas of evolutionary economic thinking can be applied to the investigation of the different domains of economics also without providing a certain and indisputable answer to the previous question. Indeed, partly following, even if not subscribing to, the Friedmanian idea that the effectiveness of a theoretical framework should be solely judged on its ability to reproduce and explain observed phenomena, one could simply start from the 'evolutionary' metaphor and see what consequences it brings to the design of economic models. As argued in Frenken and Boschma (2007), the development of an evolutionary approach to economic geography could suggest new ways of explaining the observed patterns that characterize the uneven spatial distribution of economic activities. In the spirit of the foregoing 'minimalistic' research agenda, we try to complement the bottom-up theorizing suggested there with a deeper understanding of the differences that an evolutionary inspired modeling is likely to produce with respect to more traditional approaches.

To be brought to its completion our exercise requires a twofold specification. First, we need to identify a simple formal model, based on clear assumptions, which can serve as a generic analytical framework. Second, we have to consider which hypotheses are to be put forward in order to imbue this model with the spirit of evolutionary economic geography. We address the first requirement by choosing, as a starting point, the simple two-location and multi-firm model described in Krugman (1991). This model already encompasses the idea of increasing returns and of the relevance of feedback mechanisms in shaping the aggregate economic pattern. It is well rooted in the tradition of new economic geography and, as such, constitutes a perfect benchmark for our comparative exercise. Concerning the second requirement, in line with the discussion in Boschma and Frenken (2006) and Boschma and Martin (2007), and partly inspired by the critical survey in Martin (1999), we assume the following three aspects as baseline characters of our evolutionary modeling. First, the interaction between economic agents should take place not only through market mechanisms, but also through localized, idiosyncratic interactions. Second, the flow of time should be present in the model and the decisions of economic agents, together with their consequences, should be put in an explicit time dimension. Third, the heterogeneity of firms' behavior should not be captured by a simple 'noise' term acting as a perturbation around a deterministic equilibrium. Rather, it should enter as an essential ingredient in the description of the model and in the determination of the final aggregate outcome (Granovetter, 1978; Schelling, 1978).

More precisely, we take as a starting point the model introduced in Forslid and Ottaviano (2003) and developed in Bottazzi and Dindo (2008). The latter extends Krugman (1991) by introducing a positive technological externality, assumed not tradable across locations, and by considering workers who are not mobile, which is equivalent to assuming that firms' locational decisions and the reallocation of capital goods take place over a much shorter time scale than the one characterizing work-force flows. Inside this simple economy, we consider a heterogeneous population of profit maximizing firms that independently choose where to locate their production. The model is characterized by a simple entry–exit process, and we consider a truly dynamic setting in which the locational decision of each firm is affected by the past decisions of others. As in Bottazzi et al. (2007), we assume that firms keep revising their decisions as new locational choices affect their profits. As we shall see, this updating choice process is able to generate a self-reinforcement mechanism similar to that described in Dosi and Kaniovski (1994).

The idea that localized externalities might explain agglomeration, even in the absence of workers' mobility, has been explored by several contributions inside new economic geography literature. For instance Krugman and Venables (1996) assume a vertically structured economy with localized input–output linkages, while Martin and Ottaviano (1999) consider location-specific R&D sectors that introduce different products in different locations. Baldwin and Forslid (2000) consider geographical distributions of economic activities as driven by a growth process fueled by human capital accumulation and knowledge spillovers. A drawback of these works is that, in general, they derive equilibria conditions without the complete and explicit characterization of firms' profit functions. This specification becomes however necessary when one has to design the choice procedure of firms in a dynamic environment. In order to obtain explicit expression for the profit function, we take a simpler approach: we introduce technological externalities in the form of a baseline 'cost sharing' assumption, according to which fixed production costs are shared across all firms within a given location.

The cost sharing assumption makes the model in Bottazzi and Dindo (2008) particularly suitable for the present exercise because, while remaining simple and analytically tractable, it allows for a twofold dependence of firm profits on the activity of the other firms. Using the terminology of Scitovsky (1954), this dependence takes the form of both a pecuniary externality and a technological externality. In this way, firms' profits depend on the interplay of an indirect interaction mediated by the market, which corresponds to a pecuniary externality, and a localized direct interaction, which corresponds to a technological externality. As it turns out, the former acts against the creation of production clusters while, by assumption, the latter promotes them.

Inside this framework, we analyze an explicit firms locational decision process. Our aim is to characterize the long run geographical distribution emerging from this process and relate it to the interplay of the two forms of externality. Since we explicitly introduce the time dimension in our analysis, we are also able to address history dependent phenomena. In particular, we are able to investigate how the initial state of the economy affects firms' decisions and to show that, because of firms' heterogeneity, when agglomeration occurs it is characterized by a transient nature.

This chapter is organized as follows. In section 2 we briefly describe the model and its assumptions. In section 3 we study the static setting, and derive the geographical distribution when the model is solved by assuming instantaneous equilibrium between firms choices. Starting from the previously identified equilibria, section 4 introduces both heterogeneity in agents' decisions and an explicit dynamics across time, discussing what kind of differences are observed with respect to the static case. Finally section 5 summarizes our main findings and suggests some possible further developments.

2. The model

The following model is a simplification of that described in Bottazzi and Dindo (2008), where more details can be found. Consider a two-location economy. In each location there are I households.[1] Each household is a 'local' worker, that is, he supplies labor to firms located where he resides, and a 'global' consumer, that is, he can buy goods produced in both locations and traded in a global market. The economy has two sectors: manufacturing and agriculture. In both sectors production is localized. The agricultural good is homogeneous, whereas the manufacturing good is made by differentiated products. Location $l = 1, 2$ has n_l manufacturing firms and the total number of firms is $n_1 + n_2 = N$. Without loss of generality, we assume that N is even. In order to consume manufacturing goods produced in the location where they do not reside, consumers have to pay a transportation cost $\tau \in (0, 1]$, which takes the form of an iceberg cost: for each unit of good shipped, only a fraction τ arrives at the destination. This is equivalent to saying that consumers pay a price p/τ for each unit of good they have to import. The higher the value of τ, the lower the cost of transporting the goods. Agricultural goods are traded at no costs. Agents' consumption behavior is specified by the following

Assumption 1: Each agent chooses among the agricultural good and the N different manufacturing products so as to maximize the following utility function:

$$U = C_A^{1-\mu} C_M^{\mu} \tag{24.1}$$

where the utility of bundle C_M is of constant elasticity of substitution (CES) type,

$$C_M = \left(\sum_{i=1,N} c_i^{\frac{\sigma-1}{\sigma}} \right)^{\frac{\sigma}{(\sigma-1)}} \quad \sigma > 1 \tag{24.2}$$

and each product c_i is produced by a different firm $i = 1,\ldots,N$.

Assumption 1 implies that the N products are substitutes and that σ is the mutual elasticity of substitution (see Dixit and Stiglitz, 1977). The higher σ, the more the products are substitutes and the more price differences impact on consumers' demand. Since, because of CES utility, agents value diversity, we have implicitly assumed that each firm produces a different product, so that N is both the number of manufacturing firms and the number of manufacturing products available for consumption.

The agricultural sector uses only labor as an input under constant returns to scale with unitary marginal costs. Because of the large number of potential producers, $2I(1 - \mu)$ at equilibrium, the agricultural market is perfectly competitive and the agricultural good is sold at its marginal cost.

Also manufacturing firms use only labor as input and their technologies are characterized by a common, industry-specific, marginal cost and a local, location-specific, fixed cost. Formally this leads to Assumption 2.

Assumption 2: The labour v_i that each firm $i = 1, \ldots, N$ needs to produce an amount y_i of output is given by:

$$v_i = \beta y_i + \alpha_{l_i} \tag{24.3}$$

where marginal cost β is constant across firms and across locations and the fixed costs α_{l_i} depend on the location l_i occupied by firm i.

Assumption 2 implies that we are in the presence of economies of scale, that is, an increase in output causes a decrease in each firm's average costs. Firm i profit is given by:

$$\pi_i = p_i y_i - w_i v_i = y_i(p_i - w_i\beta) - w_i\alpha_{l_i} \qquad i = 1,\ldots, N \tag{24.4}$$

where w_i is firm i cost of labor.

Before looking for markets' equilibria notice that, because of perfect competition and constant returns to scale in agricultural production, agricultural wages are equal to agricultural prices. Moreover, because of zero transportation costs for the agricultural goods, agricultural prices, and thus wages, must be the same in both locations. Given that consumers are not mobile and the economy is at an equilibrium, it should also make no difference for a worker to work in the agricultural or in the manufacturing sector. As a result wages in the two sectors, and in the two locations, are equal. For this reason it is convenient to use wages to normalizes prices in the economy and set $w_i = 1$ for all i.

In order to find equilibrium prices, quantities, profits in the manufacturing sector, and the resulting geographical distribution of firms, one should in principle analyze each of the N product markets. Nevertheless the problem can be simplified by considering only one representative market for each location. In fact, location by location, firms produce using the same technology, face the same demand (because of Assumption 1 all goods are substitutes), and the same labor supply. This implies that equilibrium prices, quantities and wages are the same for all the firms in a given location. We can thus consider only two representative product markets, one for each location l, rather than the N distinct products.

We shall compute market equilibrium prices, quantities, and profits for each fixed distribution of firms, that is, fixing n_1 and n_2. First, exploiting the CES preference structure (24.2), which gives a constant elasticity of demand, and assuming that the market structure is that of monopolistic competition, we derive firms' pricing behavior. Then using households' budget constraints we compute their total demand for the goods produced in each location, taking into account that all goods are substitutes and transportation costs impact the prices of foreign goods. Setting supply equal to demand we are able to determine equilibrium quantities and firms' profits in each location as a function of n_1 and n_2. These expressions are used, in the next section, to assess firms' geographical distribution.

Let us start from firms' pricing behavior. Consistently with our assumptions, the market structure is that of monopolistic competition, that is, each firm maximizes its profits, setting its marginal revenue equal to its marginal costs, given market demand elasticity and irrespective of other firms' behavior. Using profit function (24.4) and substituting (24.3) while setting marginal profit to zero gives:

$$p_l\left(1 + \frac{1}{\varepsilon}\right) = \beta \tag{24.5}$$

where $\varepsilon = \partial \log c/\partial \log p$ is the demand elasticity. Given Assumption 1, it holds that:

$$\varepsilon = -\sigma$$

which together with (24.5) implies:

$$p_l = \beta\frac{\sigma}{\sigma - 1} \tag{24.6}$$

Since the price does not depend on the location index, local prices are equal and it holds $p_1 = p_2$.

Denote the quantity demanded by an agent who resides in location l of a product produced in location m as d_{lm}. Each demand can be determined as a function of prices and wages using the fact that relative demands under a CES utility satisfy:

$$\frac{d_{11}}{d_{12}} = \left(\frac{p_2}{p_1\tau}\right)^\sigma \quad \text{and} \quad \frac{d_{22}}{d_{21}} = \left(\frac{p_1}{p_2\tau}\right)^\sigma \tag{24.7}$$

while agents' budget constraints give:

$$\begin{cases} \mu = n_1 d_{11} p_1 + n_2 d_{12}\dfrac{p_2}{\tau} \\ \mu = n_1 d_{21}\dfrac{p_1}{\tau} + n_2 d_{22} p_2 \end{cases} \tag{24.8}$$

where, given the Cobb-Douglas formulation in (24.1), μ is the share of agents' (unitary) income used to buy manufacturing goods. Solving for the demands, one finds:

$$d_{11} = \frac{\mu}{n_1 p_1 + n_2 p_1^{\sigma} p_2^{1-\sigma} \tau^{\sigma-1}} \qquad d_{12} = \frac{\mu \tau^{\sigma}}{n_1 p_1^{1-\sigma} p_2^{\sigma} + n_1 p_2 \tau^{\sigma-1}}$$

$$d_{22} = \frac{\mu}{n_1 p_1^{1-\sigma} p_2^{\sigma} \tau^{\sigma-1} + n_2 p_2} \qquad d_{21} = \frac{\mu \tau^{\sigma}}{n_1 p_1 \tau^{\sigma-1} + n_2 p_1^{\sigma} p_2^{1-\sigma}} \tag{24.9}$$

Equating, location by location, firms' supply and consumers' demand, gives:

$$\begin{cases} y_1 = I d_{11} + \dfrac{I d_{21}}{\tau} \\ y_2 = I d_{22} + \dfrac{I d_{12}}{\tau} \end{cases} \tag{24.10}$$

where τ discounts the demand of imported goods. Plugging (24.9) in (24.10) and using (24.6) one can solve for market equilibrium quantities:

$$\begin{cases} y_1 = \dfrac{I\mu(\sigma - 1)}{\beta\sigma}\left(\dfrac{1}{n_1 + n_2\tau^{\sigma-1}} + \dfrac{\tau^{\sigma-1}}{n_1\tau^{\sigma-1} + n_2}\right) \\ y_2 = \dfrac{I\mu(\sigma - 1)}{\beta\sigma}\left(\dfrac{1}{n_2 + n_1\tau^{\sigma-1}} + \dfrac{\tau^{\sigma-1}}{n_2\tau^{\sigma-1} + n_1}\right) \end{cases} \tag{24.11}$$

Profits in each location can now be found using the latter expression together with (24.4) and (24.6). Introducing the fraction of firms in location 1, $x = n_1/N$ (so that $n_2 = (1 - x)N$), and normalizing the level of wages to one, profits can be finally written as a function of x:

$$\begin{cases} \pi_1(x) = \dfrac{I\mu}{N\sigma}\left(\dfrac{1}{x + (1 - x)\tau^{\sigma-1}} + \dfrac{\tau^{\sigma-1}}{x\tau^{\sigma-1} + (1 - x)}\right) - \alpha_1 \\ \pi_2(x) = \dfrac{I\mu}{N\sigma}\left(\dfrac{1}{x\tau^{\sigma-1} + (1 - x)} + \dfrac{\tau^{\sigma-1}}{x + (1 - x)\tau^{\sigma-1}}\right) - \alpha_2 \end{cases} \tag{24.12}$$

Each location-specific profit function in (24.12) has a positive term proportional to the total demand for goods produced in that location and a negative term equal to the location-specific fixed costs. In turn, the total demand has a domestic component, the first term in the parentheses, and an import component, the second term in the parenthesis. Both components depend on firms' geographical distribution and transportation costs. When the transportation cost is zero, $\tau = 1$, they are equal irrespective of firms' distribution. When the transportation cost increases, the domestic component increases, as local consumers substitute foreign goods with local ones. For the same reason, the export component decreases. For any given positive transportation cost, when local firm concentration increases, the local component decreases as agents have more local goods to consume, all at the same price. In the same situation the export component increases, because foreign consumers have fewer local goods to consume and find it convenient to import more. The net effects of the transportation cost on the relative profits of the two locations are appraised in the next section. However, even without knowing this effect, it can immediately be seen that market forces make the average profits independent of transportation costs. Indeed one has the following:

Proposition 1 *Consider an economy with two locations, $l = 1, 2$, and N firms, where Assumptions 1–2 are valid. The average firm's profit $\overline{\pi}$ does not depend on transportation costs and is given by:*

$$\overline{\pi} = \frac{2I\mu}{N\sigma} - x\alpha_1 - (1 - x)\alpha_2 \qquad (24.13)$$

Proof. See Appendix.

If now one assumes that $\alpha_1 = \alpha_2$, since location-specific indexes have disappeared from any variable, only market forces are at work and our model becomes close to the one of Forslid and Ottaviano (2003). Following Bottazzi and Dindo (2008) we take a different route.

Before doing so, a last remark is necessary concerning the general equilibrium setting. In our framework, labor and goods markets are at equilibrium only when total firms' profits are zero, and only provided that the demand for labor in both locations is not higher than I. Concerning the former condition, notice that profits are already zero for the competitive agricultural firms but not necessarily so for manufacturing firms. Nevertheless it is possible to set zero profit for the manufacturing sector, imposing a long-run equilibrium condition on the size of the economy. We shall do so in Assumption 4 below. Concerning the latter condition, one has that, because of labor market segmentation (no mobility), labor demand in both locations should be lower than I. Straightforward computations show that this amounts to imposing a restriction on the preferences for manufacturing goods, namely $\mu < \sigma/(2\sigma - 1)$, which we will assume to hold from now on.[2] As a result, provided that preferences for manufacturing goods are not too strong and on imposing a long-run zero profit condition on the number of firms, prices and quantities as in (24.6) and (24.11) guarantee that both labor and good markets are at equilibrium.

Technological externalities

By retaining a dependency of the fixed cost α on the location index, we introduce a localized technological externality because of direct firms' interaction, that is, not mediated by market forces (Scitovsky, 1954).

Assumption 3: 'Cost sharing' hypothesis. Firms' fixed costs α_l decrease with the number of firms located in l according to:

$$\alpha_1 = \frac{\alpha}{2x_1}, \quad \text{where } x_l = \frac{n_l}{N}, l \in \{0, 1\} \qquad (24.14)$$

Assumption 3 represents a positive technological externality in the form of a baseline 'cost sharing': the larger the number of firms in one location, the lower the fixed costs these firms bear in the production activity. Since the fixed cost paid by firms in a given location decreases proportionally with the number of firms populating that location, the total fixed cost paid remains, location by location, constant. This effect can be thought of as an up-front cost paid to improve access to skilled labor, the more firms in one location, the smaller each firm's investment in training, or as a cost for services or infrastructure use, which is evenly shared among all the active firms in one location.

An important feature of the specific form of 'cost sharing' introduced in Assumption 3 is that it doesn't modify the total fixed costs paid by the industry. This has consequences on the computation of firms' average profit.

Corollary 1 *Consider an economy with two locations, $l = 1, 2$, and N firms, where Assumptions 1–3 are valid. Total fixed costs in each location are equal to $\alpha N/2$. The average firms' profit $\bar{\pi}$ does not depend either on the distribution of firms x or on transportation cost τ and is given by:*

$$\bar{\pi} = \frac{2I\mu}{N\sigma} - \alpha \tag{24.15}$$

Proof. See Appendix.

Before we start to look for geographical equilibria, that is, those spatial distributions of firms where they have no incentives to change location, notice that, without restrictions on the parameters' values, there could exist economies characterized by negative profits. In this case, we would expect firms to exit the economy. On the other hand, if profits were positive we could expect firms to enter the economy. As we consider the number of firms N in the model as given, if there are no barriers to entry, it seems reasonable to set the number N to a level that implies zero profits. By force of Corollary 1 this can be done also without knowing the geographical equilibrium distribution. Indeed profits at a geographical equilibrium must be equal to average profits, and average profits (24.15), because of Corollary 1, are independent on the geographical distribution x. Moreover, as we have explained at the end of the previous subsection, zero total profits are needed in order to guarantee that all markets are at an equilibrium. All together, it is enough to have the following Assumption.

Assumption 4: The number of firms N is such that profits at a geographical equilibrium are zero, that is:

$$N = \frac{2I\mu}{\sigma\alpha} \tag{24.16}$$

Even if, by construction, the previous assumption implies $\bar{\pi} = 0$, outside the geographical equilibrium profits can be both positive or negative so that their differential gives firms the incentive to relocate. Before moving to the analysis of these incentives and to the characterization of the geographical equilibria, it is useful to rewrite profits (24.12) incorporating Assumptions 3–4:

$$\begin{cases} \pi_1(x) = \dfrac{\alpha}{2}\left(\dfrac{1}{x + (1-x)\tau^{\sigma-1}} + \dfrac{\tau^{\sigma-1}}{x\tau^{\sigma-1} + (1-x)}\right) - \dfrac{\alpha}{2x} \\ \pi_2(x) = \dfrac{\alpha}{2}\left(\dfrac{1}{x\tau^{\sigma-1} + (1-x)} + \dfrac{\tau^{\sigma-1}}{x + (1-x)\tau^{\sigma-1}}\right) - \dfrac{\alpha}{2(1-x)} \end{cases} \tag{24.17}$$

Given the firms' production costs α, the products' elasticity of substitution σ and the transportation cost τ, the distribution of firms between the two locations, x, determines, through (24.17), the levels of profit. Notice that in (24.17), differently from (24.12), both the demand driven term and the fixed cost term are functions of the geographical distribution of firms. The first dependence is mediated by market forces (pecuniary externality), whereas the second dependence is brought in by the 'cost sharing' hypothesis (technological externality).

3. Geographical equlibria

In this section we investigate the static geographical equilibria of the system, that is, those distributions of firms x such that, in the search for higher profits, each firm located in 1 has no incentives to move to 2 and vice versa. Geographical equilibria can be of two types: 'border' equilibria and 'interior' equilibria. A border equilibrium occurs when firms concentrate in one location, say 1, and profits in 1 are higher than profits in 2. As all the firms are in 1, no other firm can respond to this difference in profit opportunities. Candidates for border equilibria are $x = 1$, when all firms are in 1, and $x = 0$, when all firms are in 2. Conversely, an interior equilibrium occurs when firms distribute between the two locations, that is $x \in (0, 1)$, profit levels are equal in both regions, and firms do not have any incentive to change their location. Using profits in (24.17) we will derive results for the existence and uniqueness of geographical equilibria, border and interior, for all the different parameterizations of the economy. This static[3] analysis, which owes considerably to Bottazzi and Dindo (2008) where more details can be found, is useful to understand the interplay of pecuniary and technological externalities and constitutes a useful step for the development of the evolutionary dynamic analysis of the next section.

The respective role of each externality in determining profit differentials and thus the aggregate geographical equilibrium can be judged by looking at the shape of the profit functions and by keeping in mind that, because of transportation costs, local prices are lower than foreign prices, and thus local demand impacts firms' level of output more than foreign demand. Consider the pecuniary externality term alone, for example, set $\alpha = 0$ in (24.17). For definiteness, consider profits in 1 (results for profits in 2 follow in the same way). For small x, that is, few firms in location 1 and many firms in location 2, each firm in 1 faces high local demand and low foreign demand. Because of the different impact of local and foreign demand, the level of output of firms in 1 is high and profits are high too. As x increases, the local demand for these firms decreases, so that profits decrease too. As the concentration of firms in 1 increases further, for a sufficiently large value of x, the demand coming from the consumers in 2, where very few firms are left, is more and more directed to 1 and the profits and the profits of firms located in 1 increase again. Profits are thus U-shaped, with $\pi_1(x)|_{\alpha=0}$ first decreasing and then increasing in x. Since a firm makes the most profits when alone in one location, we have $\pi_1(x = 0)|_{\alpha=0} > \pi_1(x = 1)|_{\alpha=0}$ so that the border distributions 0 and 1 are never an equilibrium. In fact, when all firm's are located in one region it is always more profitable to move to the other region. If the transportation cost is increased (decreased), the variation in profits as a function of x is more (less) pronounced but the general shape of the profit function is preserved. As a result, the overall agglomeration effect of the pecuniary externality is always 'negative', in the sense that it works against concentration of production.

The above picture changes completely when one considers also the technological externality terms introduced by the 'cost sharing' assumption, that is, $\alpha > 0$ in (24.17). The panels in Figure 24.1 show graphs of $\pi_1(x)$ and $\pi_2(x)$ in this case. Profits are given by the superposition of a monotonically increasing technological externality to the U-shaped market-driven pecuniary externality term. With low transportation costs (high

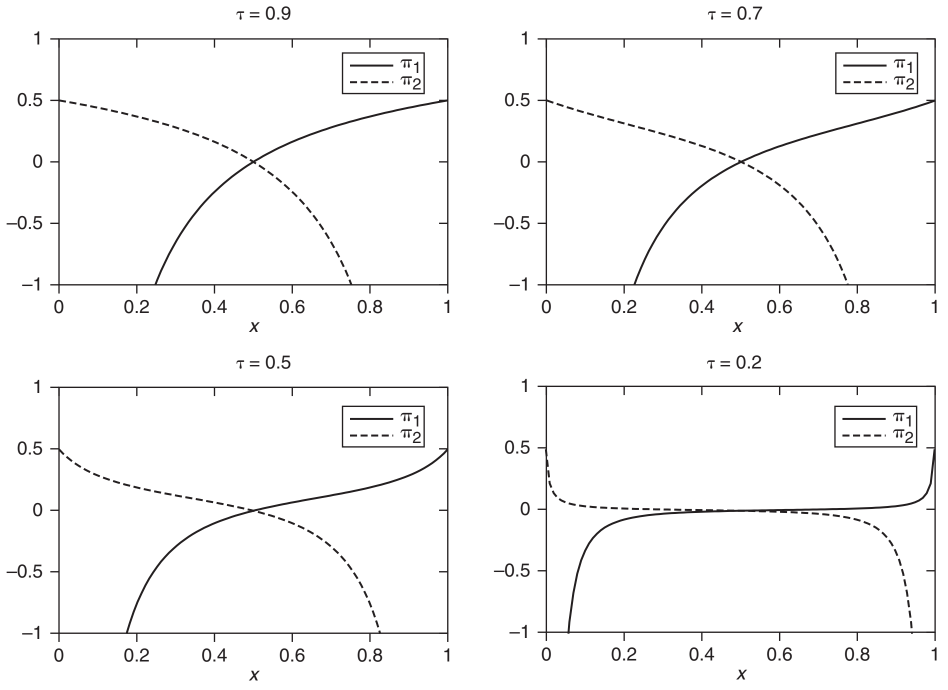


Figure 24.1 *Location profits (24.17) as a function of firms' geographical distribution for different values of the transportation cost* τ. *Other parameters are* $\sigma = 4$ *and* $\alpha = 1$

τ) the profit function is essentially determined by the 'cost sharing' term and is monotonically increasing with decreasing marginal profits (upper panels of Figure 24.1). In this case, when firms' concentration is low, firms do not benefit from the technological externality and their profits are low too, but when concentration increases, profits increase monotonically as firms exploit the 'cost sharing' opportunity. The more the firms in one location, the lower the positive contribution of an extra firm locating there, so that the marginal profit decreases. With high transportation costs (low τ) the shape of the profit function is still monotonic (bottom panels of Figure 24.1), but marginal profits are first decreasing and then increasing. With low firm concentration the technological externality dominates and marginal profits are decreasing. As the concentration increases, the positive effect of the cost sharing is almost offset by the negative market interaction, which acts as a constraint on the local demand faced by firms. In this case, even if profits are still increasing the marginal profit is almost zero. As the concentration of firms increases further, profits increase more steadily because low local demand is now compensated by the foreign demand, so that the contribution of the pecuniary externality is positive too. Judging from Figure 24.1, irrespective of the transportation costs, the positive effect of technological externalities dominates the negative effect of pecuniary externalities: firms make most profits by agglomerating on one side and border distributions are always an equilibrium. This is formalized in Proposition 2.

Proposition 2 *Consider an economy with two locations,* $l = 1, 2$ *where Assumptions 1–4 are valid. Call x the fraction of firms located in* 1. *There always exist two, and only two, geographical equilibria given by the border distribution* $x_1^* = 1$ *and* $x_0^* = 0$. *In particular, the unique distribution where profits are equal,* $x^* = 0.5$, *is never an equilibrium.*

Proof. See Appendix.

According to the previous proposition, the distribution with half of the firms located in 1 and the other half located in 2, which is the unique case where $\pi_1 = \pi_2$, is never a geographical equilibrium: even if profits are equal, incentives are such that firms move away and agglomerate. Only when all firms are located either in 1 or 2 are there no incentives to change location.

Notice that, even if transportation costs do affect the shape of each location's profit function, they have no impact in characterizing the geographical equilibria of the economy. Conversely, as we shall see in the following section, transportation costs play a major role in shaping the results of the evolutionary model, even in the long run.

4. Evolutionary firms' dynamics

In the previous section we have shown that, when the technological externality term is introduced, firms agglomerate in one of the two locations, irrespective of transportation costs τ or the relevance of technological spillover as dependent on α. This abrupt behavior would prescribe that any sector in which even a minimal level of localized non-pecuniary externalities operate should display a so-called core–periphery structure. This is clearly at odds with empirical observations. Notice that this conclusion would remain a fortiori valid if instead we had considered workers' mobility with endogenous wage setting, thus introducing a feedback effect that reduces (or inverts) the push of pecuniary externalities towards a symmetric geographical distribution. This effect ultimately reinforces the conclusion that in the presence of technological spillovers only a core–periphery structure represents an equilibrium. We end up in the uncomfortable situation of having a single possible equilibrium, implying the impossibility of performing an empirical analysis or deriving policy implications. A possible way out from this impasse, as we will show, is to extend the notion of geographical equilibrium to include an explicit dynamics describing firms' locational decisions. The foregoing analysis is, indeed, essentially static and thus silent on the results of firms' interactions out of equilibrium. As a consequence, it is not clear what happens when the initial concentration of firms is not at an equilibrium level; in particular, whether one should expect firms to agglomerate in location 1 or in location 2.

In this section we extend our analysis by introducing heterogeneity in preferences at the single firm level and by explicitly modeling firms' decisions in time, that is, by allowing for a dynamic location-specific mechanism. Suppose that the individual utility of firm i derived from locating in l_i can be written as:

$$\pi_i = \pi_{l_i} + e_{i,l_i} \tag{24.18}$$

where π_{l_i} are as in (24.17) and e_{i,l_i} represents an idiosyncratic profit component intended to capture firm-specific characteristics, like differences in productive efficiency leading to different fixed costs or individual preferences for one particular location, because of, for instance, existing social linkages. At every time step a firm is randomly chosen to exit the economy. At the same time a new firm enters and chooses whether to be located in 1 or in 2 by comparing the individual utilities in (24.18). As long as the distribution of $e_{i,l}$ across firms is well behaved (see Bottazzi and Secchi, 2007; Bottazzi et al., 2008, for details) the resulting probability of choosing l is given by:

$$\text{Prob}_l = \frac{e^{\pi_l}}{e^{\pi_1} + e^{\pi_2}}, \ l \in \{1,2\} \tag{24.19}$$

The fact that the locational choice is probabilistic derives from the assumption that the new entrant possesses preferences, or faces costs, that are not fixed, but contain an individual component that is randomly extracted from a given distribution.

When the probability of choosing location l is given by (24.19), Bottazzi and Secchi (2007) show that, if the exponentials of profits are linearly changing in the number of firms, it is possible to compute the long-run stationary distribution of the entry–exit process. Thus, to exploit this result we need a linearized version of the exponential profit functions. We can naturally obtain it as the deviation from the middle point $x^* = 0.5$, that is, the unique point where profits are equal.

Proposition 3 *Consider an economy with two locations,* $l = 1, 2$, *where Assumptions 1–4 are valid. Denote the linearization of location l exponential profits around* $x^* = 0.5$ *as* c_l, *and the number of firms in location l as* n_l. *Linearized exponential profits are given by:*

$$c_l = a + bn_l, \qquad l = 1, 2 \tag{24.20}$$

where:

$$a = 1 - \frac{4\alpha\tau^{\sigma-1}}{(1 + \tau^{\sigma-1})^2}$$

$$b = \frac{4\alpha^2\sigma\tau^{\sigma-1}}{I\mu(1 + \tau^{\sigma-1})^2} \tag{24.21}$$

Proof. See Appendix.

We shall call the term a in (24.21) the 'intrinsic profit'. This is the part of the common profit that is entirely dependent on exogenously given characteristics of the location. Conversely, the coefficient b in (24.21) captures the marginal contribution of a firm to the profit level of the location in which it resides. We shall call it the 'marginal profit'.[4] In our case, this coefficient captures the total effect of pecuniary and technological externalities. Because of the leading effect of the latter it is always positive but, because of the presence of market mediated interactions, it is dependent on transportation costs. Specifically, the marginal profit is increasing with the value of τ. When transportation costs are high (low τ) each firm's marginal contribution to the location profit is small, whereas when transportation costs are low (high τ) the marginal contribution is large.

Given the linearization in (24.20), the following proposition characterizes the long-run geographical equilibrium distribution.

Proposition 4 *Consider an economy with two locations,* $l = 1, 2$, *where Assumptions 1–4 are valid. The economy is populated by N firms, distributed according to* $n = (n_1, n_2)$. *At the beginning of each period of time a firm is randomly selected, with equal probability over the entire population, to exit the economy. Let* $m \in \{1, 2\}$ *be the location affected by this exit. After exit takes place, a new firm enters the economy and, conditional on the exit that occurred in m, has a probability:*

$$\text{Prob}_l = \frac{a + b(n_l - \delta_{l,m})}{2a + bN}$$

of choosing location l, where a and b are given by (24.1). This process admits a unique stationary distribution:

$$\pi(n) = \frac{N!C(N, a, b)}{Z(N, a, b)} \prod_{l=1}^{2} \frac{1}{n_l!} \vartheta_{n_l}(a, b) \tag{24.22}$$

where:

$$C(N, a, b) = 2a + \left(1 - \frac{1}{N}\right)bN \tag{24.23}$$

$$\vartheta_n(a, b) = \begin{cases} \prod_{h=1}^{n} [a + b(h - 1)] & n > 0 \\ 1 & n = 0 \end{cases} \tag{24.24}$$

and $Z(N, a, b)$ *is a normalization factor which depends only on the total number of firms N, and the coefficients a and b.*

Proof. See Propositions 3.1–3.4 of Bottazzi and Secchi (2007).

Figure 24.2 shows results from a simulation of the entry–exit process for two different values of the transportation cost τ. The left panel shows 50,000 iterations of the process,

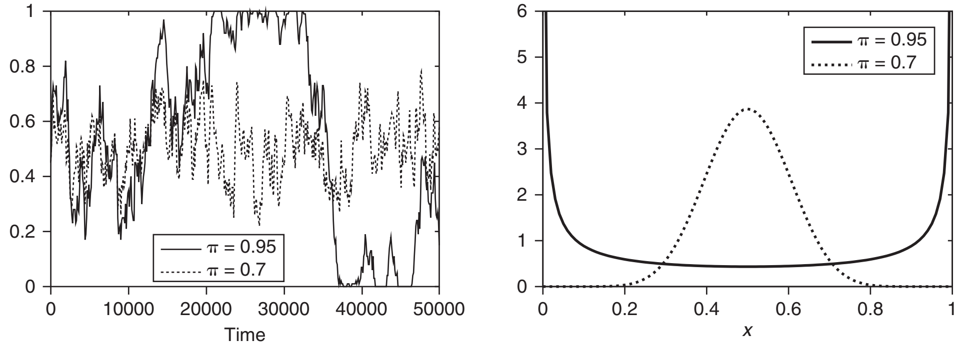


Figure 24.2 *Entry–exit process for different values of the transportation costs. Left panel: 50,000 simulations of the entry–exit process for different values of the transportation cost* τ. *Right panel: Long-run stationary distribution of the entry–exit process simulated in the left panel. In both panels the parameters are* $\sigma = 4$, $\alpha = 1$, $\mu = 0.5$ *and* $I = 800$

whereas the right panel plots the corresponding long-run distributions as characterized by Proposition 4.

With low transportation costs ($\tau = 0.95$) the long-run distribution is clustered around the two extreme values, $x = 0$ and $x = 1$, confirming the prediction of the static analysis. However, the simulation of the entry–exit process (left panel of Figure 24.2) shows that agglomeration is only a meta-stable state. One location can become much larger than the other for several time steps, like location 2 which, in the simulation shown, attracts almost all firms in the periods between 2000 and 3500, but at some point the cluster abruptly disappears and the other location can take over. This behavior is well in accordance with the bimodal nature of the equilibrium distribution (see right panel of Figure 24.2). In fact, the equilibrium distribution represents the unconditional probability of finding the system in a give state. This probability can thus be very different from the frequency with which this particular state is observed over a finite time window.

Conversely, for higher transportation costs (τ =0.7), agglomeration is 'almost' never observed: firms spatial distribution is now fluctuating around x^* =0.5 (see left panel of Figure 24.2). Even if the static analysis predicts agglomeration, the equilibrium distribution of the stochastic system, reported in the right panel of Figure 24.2, shows that the most likely geographical distribution has an equal number of firms per location, irrespective of the fact that the point $x^* = 0.5$ is never a static geographical equilibrium. In general one has the following:

Proposition 5 *Consider the entry and exit process described in Proposition 4. When the marginal profit is bigger than the intrinsic profit, $b > a$, the stationary distribution (24.22) is bimodal with modes in $x = 0$ and $x = 1$; when $b < a$ the stationary distribution is unimodal with mode in $x = 0.5$; and when $a = b$ the stationary distribution is uniform.*

Proof. See Appendix.

Given our dynamic locational decision process, the previous proposition clarifies that the shape of the geographical equilibrium distribution does ultimately depend on the relative size of the marginal profit b and the intrinsic profit a. When marginal profits are bigger than intrinsic profits the distribution has mass on the borders of the [0, 1] interval. When marginal profits are lower than intrinsic profits the distribution has higher mass in the middle of this interval, and when they are equal every value of the geographical distribution is as likely.

Rewriting the relation $b \gtrless a$ in Proposition 5 using the definitions of a and b in (24.21), it is straightforward to derive the conditions for the unimodality or bimodality of (24.22) in terms of the values of τ, I, μ, and α:

$$4\alpha\left(1 + \frac{\alpha\sigma}{I\mu}\right) \gtrless \frac{(1 + \tau^{\sigma-1})^2}{\tau^{\sigma-1}}$$

The left panels of Figures 24.3 and 24.4 have been obtained using the latter inequality: they show which distributional shape is observed in the different regions of the plane (I, τ) and (α, τ) respectively. In the white area agglomeration is most likely (bimodal distribution), whereas in the dark area equidistribution is most likely (unimodal distribution). In the right panels the stationary distributions computed at the corresponding

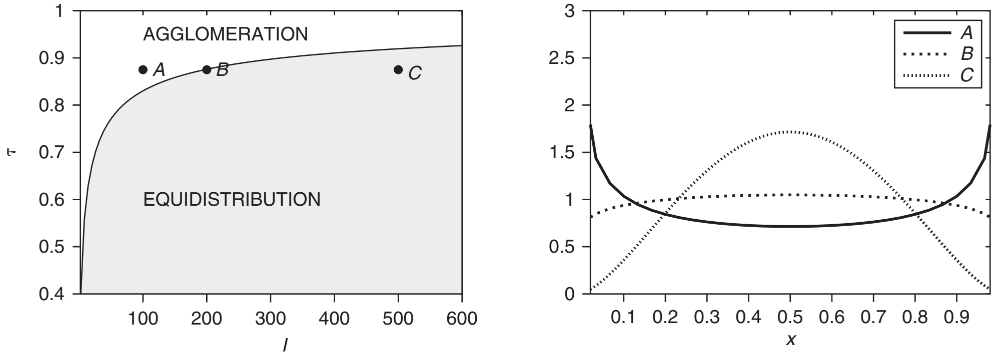


Figure 24.3 *Entry–exit process for different values of the number of residents **I**. Left panel: A portion of the space (I, τ) has been divided into the 'agglomeration' area (white) and the 'equidistribution' area (shaded) according to Proposition 5. Right panel: Stationary geographical distributions computed at the points A, B, and C. Other parameters are α = 1, μ =0.5 and σ = 4, whereas N is fixed by Assumption 4*

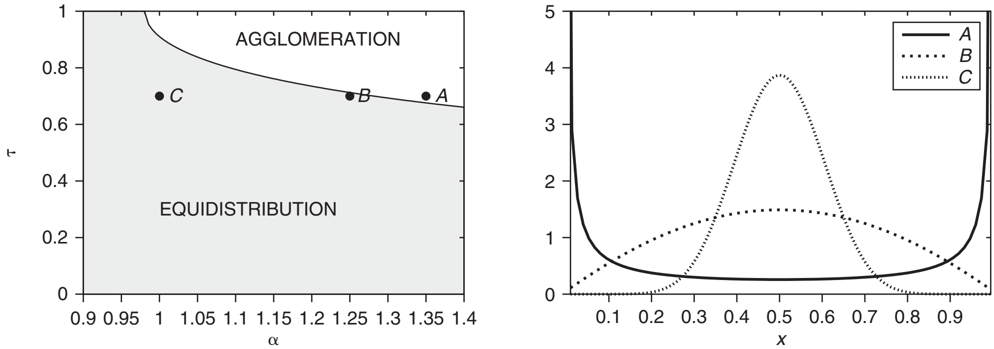


Figure 24.4 *Entry–exit process for different values of the fixed costs α. Left panel: A portion of the space (α, τ) has been divided into the 'agglomeration' area (white) and the 'equidistribution' area (shaded) according to Proposition 5. Right panel: Stationary geographical distributions computed at the points A, B, and C. Other parameters are I = 400, μ =0.5 and σ = 4 whereas N is fixed by Assumption 4*

points *A*, *B*, and *C* are shown. In both figures these points have been obtained by keeping τ fixed.

The right panel of Figure 24.3 shows that moving from small to large values of *I* while keeping τ fixed, the long-run distribution changes from bimodal to unimodal. This is because an increase in the number of residents *I* leads to a decrease in the marginal profit *b* (see 24.21). In fact, as a result of Assumption 4, the more residents there are, the more firms there are and the smaller the contribution of each firm's locational decision to the profits of other firms, that is, the smaller the marginal profit. Changing *I* corresponds to a sort of 'size' effect: increasing the size of the economy lowers the externalities so that, because of the entry–exit process, the likelihood of observing agglomeration is lowered.

The right panel of Figure 24.4 shows that, keeping τ fixed, an increase in the fixed cost parameter α leads, in general, to more agglomerated economies. It is so because an increase in α decreases the intrinsic profit a while increasing the marginal profit b. More precisely, α determines the scale of the profit differentials. Indeed, on the one hand, the difference between the maximum and the minimum profit is proportional to α and, on the other hand, because of Assumption 4, the higher α the lower N, so that a bigger profit difference is caused by a lower number of firms. As a result, increasing fixed costs decreases the profit each firm earns irrespective of the presence of other firms, and increases the effect of each locational choice on the profits of others. Both effects go in the direction of increasing the likelihood of agglomeration. This is a sort of 'scale' effect where increasing the scale of profits increases the likelihood of agglomeration.

Concerning the effect of the transportation cost on the shape of the equilibrium distribution, notice that the expression $\tau^{\sigma-1}/(1 + \tau^{\sigma-1})^2$, which appears in (24.21) for both a and b, but with a different sign, is an increasing function of τ. Thus, increasing the value of τ leads to an increase in the marginal profit b and a decrease in the intrinsic profit a. This means that low transportation costs, that is, high values of τ , favor agglomeration, while high transportation costs favor equidistribution. Indeed, when transportation costs are low, the pecuniary externality is relatively weak and the technological externality relatively strong. In terms of the entry–exit process, the choice of a firm to relocate its activity has an high impact on the level of profits. Consequently, it is likely to trigger other relocations and, eventually, a strong agglomeration is observed. Conversely, when transportation costs are high, the pecuniary and technological externalities almost offset each other. This implies that marginal profits are small and intrinsic profits dominate, so that each locational choice has a very small impact on the general level of profits. The attracting force of each location does not depend on the externality term and, given the symmetry of the two locations, equidistribution is likely to be observed.

5. Conclusion

We have analyzed a model of firms' location in geographical space where firms interact both indirectly, through market interactions, and directly, through technological externalities, and where workers are not mobile. In this simple framework we have briefly discussed the general equilibrium static case, identifying the possible geographical equilibria, that is, the spatial distributions in which firms do not have any incentive to relocate their activities. We have showed that in this case the 'cost sharing' assumption implies long-run agglomeration, irrespective of the number of consumers, their preferences, and transportation costs. Then we have extended the analysis to include heterogeneity in firms' preferences and an explicit time dynamics in their choices, thus obtaining a stochastic model of firms' dynamics. We have been able to characterize the long-run geographical distribution of the process for different specifications of the economy. This analysis has revealed that, contrary to the static equilibrium analysis, when an explicit entry–exit dynamics is assumed to characterize the locational decision of firms, the economy can evolve towards two different long-run scenarios. In the first scenario, where externalities are stronger than intrinsic location profits, which typically occurs for low transportation costs, the long-run geographical distribution is bimodal with modes at the extremal outcomes $x = 0$ and $x = 1$. Agglomeration is thus the most likely event but, as simulations show, this does not mean that once agglomeration on one side has been achieved, the situation is stable. In fact, turning points exist where the mass of firms moves from one location to the other. In the second scenario the long-run geographical distribution has a unique mode. In this case, the most likely occurrence is having half of the firms located in one region and the other half in the other region. However, because of the stochastic nature of the process, fluctuations around this average level are present. This scenario is typically associated with high transportation costs, and occurs, in general, when the effect of externalities is weak with respect to the intrinsic profit levels of each location.

Summarizing, the main contribution of the foregoing analysis is to show how firms' heterogeneity and an individual choice process act as brakes or constraints on firms' agglomeration, even when strong incentives to locate in already populous locations exist. Moreover, having introduced an explicit time dimension, we have given history a role. Indeed the time dimension matters in two respects: first, the initial distribution of activities across two locations does influence the subsequent observed distributions and, second, when agglomeration is observed, because of stochastic fluctuations, it is only a metastable phenomenon. That is, by waiting long enough, the cluster eventually disappears, just to be recreated soon after, with probability 1/2, in the other location.

Our model can be extended in several directions. First of all, the 'cost sharing' assumption, while useful, is admittedly ad hoc. More careful modeling is probably needed. The effort should not be restricted to the notion of technological and/or knowledge spillover, which might even be characterized by a pecuniary nature, see, for example, Antonelli (Chapter 7, this volume), but could encompass also other, possibly negative, sources of interactions that are not market mediated, like pollution and/or congestion effects. A second extension of the model would be to generalize consumers' behavior along the same lines we followed to describe firms' behavior. Whereas in the present version of the model consumers are homogeneous and maximize the same CES utility function, it would be interesting to assume that consumers are heterogeneous and to explicitly model their consumption decision in time. In that case, changing the size of the economy would imply, because of varying idiosyncrasies in consumers' demand, a change in the amplitude of profit fluctuations. This, in turn, would impact the likelihood of observing agglomerated outcomes, probably reducing it.

In any case, we are aware that the ultimate test will be to confront our model with real data. An interesting aspect of the discrete choice model we implemented is that it leads quite easily to empirical applications. An exercise in this direction has already been performed in Bottazzi et al. (2008), where the parameters characterizing the geographical equilibrium distribution have been estimated in several sectors of the Italian manufacturing industry. The present work moves in the direction of developing a theoretical framework able to provide deeper and more informative economic interpretations of these econometric exercises.

Notes

* Thanks to the participants to the DIMETIC Summer School 'Geography of Innovation and Growth', Pecs, Hungary; the First DIME Scientific Conference, Strasbourg, France; and seminars given at the Università degli Studi di Napoli 'Parthenope' and at the Scuola Superiore S'Anna, Pisa, Italy. The research

that has led to this work has been supported by the EU FP6 STREP Project CO3 Common Complex Collective Phenomena in Statistical Mechanics, Society, Economics, and Biology and by the European Union NoE DIME. All usual disclaimers apply.

1. The generalization to different numbers of agents in each location will be considered in future work. Preliminary analysis shows that it doesn't significantly modify the results.
2. This is the same condition found in Forslid and Ottaviano (2003). See Bottazzi and Dindo (2008) for more details.
3. Technically our geographical equilibrium corresponds to a Nash equilibrium in pure strategies of the one shot game where each firm in a group of N has to choose whether to be located in 1 or in 2 and payoffs are given by profits.
4. In the terminology of Bottazzi and Secchi (2007) the intrinsic profit corresponds to the location's 'intrinsic attractiveness', whereas the marginal profit is the 'social externality'.

References

Alchian, A. (1950), 'Uncertainty, evolution, and economic theory', *The Journal of Political Economy*, **58**, 211–21.

Baldwin, R. and R. Forslid (2000), 'The core–periphery model and endogenous growth: stabilizing and destabilizing integration', *Economica*, **67**, 307–24.

Boschma, R. and K. Frenken (2006), 'Why is economic geography not an evolutionary science? Towards an evolutionary economic geography', *Journal of Economic Geography*, **6**, 273–302.

Boschma, R. and R. Martin (2007), 'Editorial: Constructing an evolutionary economic geography', *Journal of Economic Geography*, **7**, 537–48.

Bottazzi, G. and A. Secchi (2007), 'Repeated choices under dynamic externalities', LEM Working Paper 2007–08, Sant'Anna School for Advanced Studies, Pisa.

Bottazzi, M. and P. Dindo (2008), 'Localized technological externalities and the geographical distribution of firms', LEM Working Paper 2008–11, Sant'Anna School of Advanced Studies.

Bottazzi, G., G. Dosi, G. Fagiolo and A. Secchi (2007), 'Modeling industrial evolution in geographical space', *Journal of Economic Geography*, **7**, 651–72.

Bottazzi, G., G. Dosi, G. Fagiolo and A. Secchi (2008), 'Sectoral and geographical specificities in the spatial structure of economic activities', *Structural Change and Economic Dynamics*, **19**, 189–2002.

Dixit, A. and J. Stiglitz (1977), 'Monopolistic competition and optimum product diversity', *American Economic Review*, **67** (3), 297–308.

Dosi, G. (1988), 'Sources, procedures, and microeconomic effects of innovation', *Journal of Economic Literature*, **26**, 1120–71.

Dosi, G. and Y. Kaniovski (1994), 'On "badly behaved" dynamics', *Journal of Evolutionary Economics*, **4**, 93–123.

Forslid, R. and G. Ottaviano (2003), 'An analytically solvable core–periphery model', *Journal of Economic Geography*, **3**, 229–40.

Freeman, C. (1986), *The Economics of Industrial Innovation*, 2nd edn, Cambridge, MA: MIT Press.

Frenken, K. and R. Boschma (2007), 'A theoretical framework for evolutionary economic geography: industrial dynamics and urban growth as branching process', *Journal of Economic Geography*, **7**, 635–49.

Granovetter, M. (1978), 'Threshold models of collective behavior', *The American Journal of Sociology*, **83** (6), 1420–43.

Krugman, P. (1991), 'Increasing returns and economic geography', *The Journal of Political Economy*, **99**, 483–99.

Krugman, P. and A. Venables (1996), 'Globalization and the inequality of nations', *Quarterly Journal of Economics*, **110**, 857–80.

Martin, P. and G. Ottaviano (1999), 'Growing locations: industry locations in a model of endogenous growth', *European Economic Review*, **43**, 281–302.

Martin, R. (1999), 'The new "geographical turn" in economics: some critical reflections', *Cambridge Journal of Economics*, **23**, 65–91.

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

Schelling, T.C. (1978), *Micromotives and Macrobehavior*, New York: W.W. Norton.

Scitovsky, T. (1954), 'Two concepts of external economies', *The Journal of Political Economy*, **62**, 143–51.

Appendix

Proof of Proposition 1 and Corollary 1
Average profits follow in a straightforward way by computing $\bar{\pi} = x\pi_1(x) + (1 - x)\pi_2(x)$ with $\pi_1(x)$ and $\pi_2(x)$ as in (24.12) and in (24.17). Total costs in each location are equal because (24.14) implies:

$$\sum_{i=1,n_1} \alpha_i = \sum_{j=1,n_2} \alpha_j = \frac{N\alpha}{2}$$

Proof of Proposition 2
First, the unique distribution of firms where profits are equal is $x = 0.5$. Indeed taking profits as in (24.17), setting $\pi_1(x)$ equal to $\pi_2(x)$, gives a first order equation in x whose unique solution is $x = 0.5$.

Geographical equilibria are pure strategy Nash equilibria (PSNE) of the one stage game where each firm in a group of N (N even) has to choose to be located in $l = 1$ or $l = 2$ and profits are given by (24.17). Denote the firm $i = 1, \ldots, N$ strategy as s_i. Firm i can choose to locate either in 1, $s_i = 1$, or in 2, $s_i = 0$. The strategy space has thus 2^N elements. A strategy profile will be denoted as s while s_{-i} will denote the strategy profile of the $N - 1$ firms apart from i. Define also:

$$x(s) = \frac{\sum_{i=1,\ldots,N} s_i}{N}$$

To complete the formalization of the game we have to specify each firm payoff for any strategy profile s. When $s_i = 1$, firm i payoff π_i is given by:

$$\pi_i(1, s_{-i}) \equiv \pi_1(x(s))$$

where $\pi_1(x)$ is as in (24.17) and $x(s)$ is defined above. When $s_i = 0$, firm i payoff is:

$$\pi_i(0, s_{-i}) \equiv \pi_2(x(s))$$

where $\pi_2(x)$ is again from (24.17). To give an example, if all firms choose location $l = 1$, so that $x = 1$, one has $\pi_i = \pi_1(1)$ for all $i = 1, \ldots, N$. If, instead, half of the firms are located in $l = 1$ and the other half in $l = 2$, so that $x = 0.5$, one has $\pi_i = \pi_1(0.5)$ if $s_i = 1$ and $\pi_i = \pi_2(0.5)$ otherwise. A strategy profile s^* is a PSNE if and only if:

$$\pi_i(s_i^*, s_{-i}^*) \geq \pi_i(s_i, s_{-i}^*) \text{ for all } s_i = 0, 1,\ i = 1, \ldots, N \tag{24A.1}$$

The only candidates to be PSNE are those strategy profiles s for which $x(s) \in \{0, 1, 0.5\}$. We start by showing that every s^* such that $x(s^*) = 1$ or $x(s^*) = 0$ is a PSNE profile. From (24.17) it holds that $\pi_1(1) > \pi_2(x)$ for all $x \in (0, 1)$ and $\pi_2(0) > \pi_1(x)$ for all $x \in (0,1)$, so that (24A.1) is satisfied. Then consider a strategy profile s^* such that $x(s^*) = x^* = 0.5$. Given (24A.1), it is a PSNE if and only if:

$$\pi_1(0.5) \geq \pi_2\left(0.5 - \frac{1}{N}\right) \tag{24A.2}$$

$$\pi_2(0.5) \geq \pi_1\left(0.5 + \frac{1}{N}\right) \tag{24A.3}$$

Since the two locations are identical and x is the share of firms choosing location 1, by symmetry it holds that $\pi_1(x) = \pi_2(1 - x)$. Using this relation we can rewrite both (24A.2) and (24A.3) as:

$$\pi_1(0.5) \geq \pi_1\left(0.5 + \frac{1}{N}\right)$$

The latter is satisfied if and only if the function $\pi_1(x)$ is not increasing at $x = 0.5$. Direct computation of $d\pi_1(x)/dx|_{x=0.5}$ (see the following proofs for the explicit expression) shows that this is never the case, implying that the symmetric distribution is never a PSNE.

Proof of Proposition 3

Consider the Taylor expansion up to the first order of each term in (24.17) as a function of $z = x - 0.5$:

$$\pi_l(z) = \pi_l(0.5) + z(-)^l\left(2\alpha\left(\frac{1 - t^{\sigma-1}}{1 + t^{\sigma-1}}\right)^2 - 2\alpha\right) + O(z^2),\ l = 1, 2$$

Using the expressions above to linearize expπ_l in (24.19), and writing them in terms of the number of firms n_l, l =1, 2 we obtain the expressions of the linearized exponential payoff c_l,

$$c_1 = 1 - 2\alpha\left(\frac{(1 - t^{\sigma-1})^2}{(1 + t^{\sigma-1})^2} - 1\right)\left(\frac{n_1}{N} - \frac{1}{2}\right)$$

$$c_2 = 1 + 2\alpha\left(\frac{(1 - t^{\sigma-1})^2}{(1 + t^{\sigma-1})^2} - 1\right)\left(\frac{1}{2} - \frac{n_2}{N}\right)$$

This shows that a and b are given by:

$$a = 1 + \alpha\left(\frac{(1 - t^{\sigma-1})^2}{(1 + t^{\sigma-1})^2} - 1\right)$$

$$b = -\frac{2\alpha}{N}\left(\frac{(1 - t^{\sigma-1})^2}{(1 + t^{\sigma-1})^2} - 1\right)$$

which, using Assumption 4 to eliminate N, correspond to (24.21).

Proof of Proposition 5

From (24.22) it follows that the distribution is symmetric around $N/2$, that is $\pi(N/2 + n) = \pi(N/2 - n)$ for every $n = 0, \ldots, N/2$. Consequently it suffices to analyse the set $\{0, \ldots, N/2\}$.

When $a = b$, $\theta_n(a, b)$ reduces to $n!$. As a result the probability density becomes:

$$\pi(n) = \frac{N!C(N, a, b)}{Z(N, a, b)} \quad \forall n,$$

so that the distribution is uniform. The rest of the lemma will be proved by induction. First consider:

$$\begin{aligned}\pi(0) &\gtrless \pi(1)\\ \frac{N!C(N,a,b)}{Z(N,a,b)}\frac{1}{N!}\theta_N(a,b) &\gtrless \frac{N!C(N,a,b)}{Z(N,a,b)}\frac{1}{(N-1)!}\theta_1(a,b)\theta_{N-1}(a,b)\\ a(a+b)\ldots(a+b(N-1)) &\gtrless Na^2(a+b)\ldots(a+b(N-2))\\ b &\gtrless a \end{aligned} \tag{24A.4}$$

Second consider:

$$\begin{aligned}\pi(n) &\gtrless \pi(n+1)\\ &\frac{N!C(N,a,b)}{Z(N,a,b)}\frac{1}{n!(N-n)!}\theta_n(a,b)\theta_{N-n}(a,b)\\ &\gtrless \frac{N!C(N,a,b)}{Z(N,a,b)}\frac{1}{(n+1)!(N-n-1)!}\theta_{n+1}(a,b)\theta_{N-n-1}(a,b)\\ &(n+1)a(a+b)\ldots(a+b(n-1))a(a+b)\ldots(a+b(N-n-1))\\ &\gtrless (n+1)a(a+b)\ldots(a+bn)a(a+b)\ldots(a+b(N-n-2))\\ b(N-2n-1) &\gtrless B(N-2n-1)\\ b &\gtrless a\end{aligned} \tag{24A.5}$$

where the last step requires $n \leq N/2 - 1$, which is our case. From 24A.4 and 24A.5 it follows that when $b > a$ the maximum is in $\pi(0)$ (and by symmetry in $\pi(N)$), whereas, when $b < a$, the maximum is in $\pi(N/2)$.

Index