Economic Complexity and Evolution

Guido Buenstorf · Uwe Cantner
Horst Hanusch · Michael Hutter
Hans-Walter Lorenz · Fritz Rahmeyer
Editors

The Two Sides of Innovation

Creation and Destruction in the Evolution of Capitalist Economies

Springer

Economic Complexity and Evolution

Series Editors:

For further volumes:
http://www.springer.com/series/11583

Guido Buenstorf • Uwe Cantner
Horst Hanusch • Michael Hutter
Hans-Walter Lorenz • Fritz Rahmeyer
Editors

The Two Sides of Innovation

Creation and Destruction in the Evolution of Capitalist Economies

Springer

Editors
Guido Buenstorf
University of Kassel
Kassel, Germany

Uwe Cantner
Friedrich-Schiller-University Jena
Jena, Germany

Horst Hanusch
University of Augsburg
Augsburg, Germany

Michael Hutter
Social Science Research Center Berlin
Berlin, Germany

Hans-Walter Lorenz
Friedrich-Schiller-University Jena
Jena, Germany

Fritz Rahmeyer
University of Augsburg
Augsburg, Germany

ISBN 978-3-319-01495-1 ISBN 978-3-319-01496-8 (eBook)
DOI 10.1007/978-3-319-01496-8
Springer Cham Heidelberg New York Dordrecht London

Library of Congress Control Number: 2013954971

Printed on acid-free paper

Springer is part of Springer Science+Business Media (www.springer.com)

Contents

Editorial: The Two Sides of Innovation

Guido Buenstorf, Uwe Cantner, Horst Hanusch, Michael Hutter, Hans-Walter Lorenz, and Fritz Rahmeyer

Schumpeter's notion of creative destruction is certainly one of the most-cited expressions ever coined by an economist:

> *"The opening up of new markets, foreign or domestic, and the organizational development from the craft shop and factory to such concerns as U.S. Steel illustrate the same process of industrial mutation [. . .] that incessantly revolutionizes the economic structure from within, incessantly destroying the old one, incessantly creating a new one. This process of Creative Destruction is the essential fact about capitalism. It is what capitalism consists in and what every capitalist concern has got to live in."* Schumpeter (1994, [1942], pp. 82–83).

For Schumpeter, there was no doubt that innovation—coming *from within*—has a destructive side. Where there are pioneering entrepreneurs, there will also be laggards, "mere managers" (in Schumpeter's terms) who cannot keep up with the pace of change brought about by new combinations in products, processes and organizations. Product innovations may lead to the emergence of entirely new industries—and at the same time trigger the demise of existing ones the products of which are replaced by the innovation. Schumpeter also realized that the structural changes thus brought about could alter the prospects of industries, entire regions

G. Buenstorf
University of Kassel, Kassel, Germany
e-mail: buenstorf@uni-kassel.de

U. Cantner (✉)
Friedrich-Schiller-University Jena, Jena, Germany
e-mail: uwe.cantner@uni-jena.de

H. Hanusch • F. Rahmeyer
University of Augsburg, Augsburg, Germany

M. Hutter
Social Science Research Center Berlin, Berlin, Germany

H.-W. Lorenz
Friedrich-Schiller-University Jena, Jena, Germany

G. Buenstorf et al. (eds.), *The Two Sides of Innovation*, Economic Complexity and Evolution, DOI 10.1007/978-3-319-01496-8_1,

and even countries. Tellingly, he did not speak of "destructive creation" but of "creative destruction"—creativity is the qualifying attribute he assigned to an innovative process, while destruction is at its core. It is therefore only consequent that Schumpeter expected capitalism itself to be undermined by the destructive creativity of (social) innovation in the long run.

All this is, of course, well-known to evolutionary economists, and much of it has informed key contributions to evolutionary economics. Most of us will agree that capitalism is a restless process (Metcalfe 2008) and that equilibria in markets should be understood as (at most) short-living times of temporary stability during phases of rapid structural change. At the same time, it seems fair to argue that evolutionary economics has been somewhat one-sided in its treatment of innovation, with the destructive part of the innovative creation process having received relatively little attention. Adverse effects of innovation are rarely discussed, perhaps due to the underdeveloped nature of evolutionary welfare economics. Entrepreneurship and new firm formation are much more prominent topics in evolutionary economics than is the demise of companies. The work on industry life cycles has focused on explaining the shakeout in the number of firms during the growth phase of an industry, while the later demise of obsolete industries has mostly been eclipsed. Related work has looked at the fate of diversifiers into new industries, while little is known about the post-exit activities of firms withdrawing from a specific market. Many other examples could be provided for the tendency to ignore the destructive side of innovation in the thrust of research in evolutionary economics. Perhaps most striking is the disregard of Schumpeter's conjecture that capitalism itself is subject to a creative evolutionary process which may ultimately be destructive—a conjecture that resonates with the debate of the future of market economies rekindled by the recent financial and fiscal crises.

This volume offers a collection of papers devoted to innovation and its consequences. These papers were presented and discussed at three workshops in Augsburg (2008), Jena (2009) and Linz (2010) organized by the section "Evolutionary Economics" of the German Economic Association (Verein für Socialpolitik). The common theme of the three workshops was the role of innovation—with its two sides, namely creation and destruction—in the evolution of capitalist economies and also the evolution of capitalism itself. This workshop agenda is reflected by the selection of papers included in this volume.

A first set of papers looks at innovation and its effects on economic performance, addressing issues of motives, behavioral rules under uncertainty, actor properties, and technology characteristics. A second group of papers concentrates on potential consequences of innovative activities, in particular, structural change, the "innovation-mediated" effect of skill-oriented policies on regional performance, the destructive effects of innovation activities, and the question whether novelty is always good. The role of innovation in the evolution of capitalism itself is discussed in a third group of papers.

1 Innovation: Conditions to Successfully Create Novelty for Economic Development

In her paper "Agents of change", Caroline Gerschlager (WU Vienna) suggests that the theory of agents of change, which currently rests mainly on limited cognitive capabilities, be complemented by an analysis of agents' motives. Here, the capacity of actors to change their motives can be seen as another source of innovation in the sense that agents are not only affected by their ideas, institutions and technologies (as in the cognitive approach), but also deliberately affect and choose their ideas, institutions and technologies. Behavioral rules under uncertainty, as encountered in basic science, are the concern of Thomas Grebel (Friedrich Schiller University Jena; meanwhile Technical University of Ilmenau) in his paper "Network evolution in basic science". Based on an empirical study in cardiology, he discusses specific behavioral rules in the initial phase of starting a research project based on received information or perceived competition, as well as rules relating to cooperation and imitation. Applying these rules in a percolation model, his simulation analyses find that, along with the diffusion of new information, specific leader-follower as well as cooperation non- cooperation structures emerge.

Looking at inventor networks from an empirical point of view, the development of those networks and the relationship to the degree of generality of the pursued technology are analyzed by Holger Graf (Friedrich Schiller University Jena) in his paper "Inventor networks in emerging key technologies: information technology vs. semiconductors". Technologies that move to the core of the knowledge base—such as semiconductors—are considered to be key technologies allowing for further knowledge-widening activities; hence, associated structures of the inventor network should exhibit a high degree of connectedness. In contrast, other technologies which do not move very much in the knowledge space—such as information technologies—are characterized by knowledge-deepening activities, so that the emerging structure of an inventor network shows lower connectedness.

Turning to the commercialization of inventions, the problem of the transfer of the results in basic science, i.e. academic inventions, into economic usage is addressed in "Not invented here: technology licensing, knowledge transfer and innovation based on public research" by Guido Buenstorf and Matthias Geissler (University of Kassel). Licensing to external actors and licensing to academic spin-offs are alternative channels for successfully commercializing academic inventions. Their effectiveness is analyzed on the basis of data from the Max Planck Society in Germany. The results indicate that academic spinoffs, compared to external licensing, do not have a systematic disadvantage in commercializing inventions. Hence, from a policy perspective, spin-offs should not be considered an inferior solution applicable only when external licensees are not available.

In their paper "Innovation in the Age of the Fuggers", Rolf Walter and Maximilian Kalus (Friedrich Schiller University Jena) use a sixteenth century historical example to illustrate the importance of human capital for sustained firm competitiveness. (In this respect, capitalism may not have changed that much in the past

500 years.) As Walter and Kalus show, the Fugger family excelled at introducing various forms of innovation—such as new products, improved logistics, and also social innovations. Their firm was one of the most internationally diversified players of the early modern era, An outstanding capability to combine different operating areas, a high degree of internationality, a dense network of agents and well-trained personnel (excellent human capital), a most efficient logistics system and close personal contacts to the European courts and to curia and pope were some of the key elements of its sustained success. Interestingly, as true agents of change, the Fuggers transformed large parts of their enormous accumulated wealth to cultural and social capital; this resulted in social innovations such as the Fuggerei—the world's first low-income housing project—in the city of Augsburg.

2 Innovation: Induced Structural Change, Coping and Normative Assessment

A simulation study looking at innovation activities, the diffusion of innovations and the resulting structural change is presented by Frank Beckenbach, Maria Daskalakis and David Hofmann (University of Kassel) in "Agent-based modeling of novelty creating behavior and sectoral growth effects—Linking the creative and the destructive side of innovation". The multilevel model represents the relationship between agents' novelty-creating activities, on the one hand, and the growth of economic aggregates, on the other. The model reproduces some stylized facts of the knowledge-generating process, such as the persistence of actor heterogeneity, the emergence of innovating agents as a specific property of the model, and the inverse cyclical pattern of innovation and imitation.

In "Creative production in the creative industries" Michael Hutter (WZB Berlin) addresses an economy's capacity to appreciate and depreciate new contributions. Accompanying the process of creative destruction, this capacity counteracts homogenization tendencies that, according to Schumpeter, would threaten to stifle the innovative process. This dynamic becomes most evident in those segments of the economy that depend strongest on a constant flow of novelty, the cultural and the creative sector. There, most of the goods offered are information goods that generate affective sensations in their users and that are valuated via communication among amateurs and experts. The interplay between surprising new information and devices of valuation frames the production process in those industries and is decisive for its sustained growth.

A further dimension of (creative) destruction is taken up by Lars Feld (Albert Ludwigs University Freiburg & Walter Eucken Institute), Jan Schnellenbach (Ruprecht Karls University Heidelberg) and Thushyanthan Baskaran (Georg August University Göttingen) in "Creative destruction and fiscal institutions: a longrun case study of three regions". Analyzing the cases of the declining steel and mining industry in three jurisdictions, the authors investigate the distinct regional fiscal constitutions and the associated transfer payments for the obstruction or the encouragement of structural change in the private sector. For a region with fiscal autonomy, the authors find a relatively faster decline of employment in the respective sectors. For non-autonomous regions, transfers do not accelerate structural change, but have a more preserving effect.

The economic performance of regions as measured by total factor productivity, the importance of the quality of capital and of skilled labor therein, and the role of labor market policies are discussed in “Labor market integration policies and the convergence of regions: the role of skills and technology diffusion” by Herbert Dawid, Simon Gemkow, Philipp Harting (all University of Bielefeld) and Michael Neugart (Free University of Bozen). The authors contrast alternative labor market integration policies resulting in distinct regional distributions of specific skills. As a consequence, a trade-off between convergence and the level of output is identified. In the case of a closed labor market, a high degree of convergence is combined with a low level of output, whereas a more open labor market implies a higher output level combined with a lower tendency for convergence.

“Is novelty always a good thing? Towards an evolutionary welfare economics” is the title of the paper by Christian Schubert (Max Planck Institute of Economics, Jena) that concludes the second part of the volume. In this paper, Schubert questions the position that innovation is always welfare-increasing—a position usually, and mostly implicitly, adopted by policy makers trying to foster innovation activities. In discussing this issue, and by critically reflecting on traditional welfare economics, the author develops a concept of welfare in a world of variable and often incoherent preferences. In his new conception, welfare at the individual level is understood as the capacity and motivation to engage in the ongoing learning of instrumentally effective preferences. In this sense, innovation is beneficial if it enlarges a person’s freedom to choose among different paths of preference learning. Innovation policy is beneficial if it promotes an individual’s ability to learn and to explore new preferences.

3 Innovation and the Evolution of Capitalism

As became evident in the recent crises, the financial sector is a key driver of evolutionary change and creative destruction in modern capitalist economies. In “Dynamic Circular Flow Models with Innovations—Some methodological aspects of the banking and finance crisis from the point of view of evolutionary economics”, Adolf Wagner (Leipzig University) focuses on the importance of money and finance for aggregate macroeconomic stability over intermediate periods of time. Referring to the concept of circulation disequilibrium, he highlights the role of institutional transactors in managing stocks and flows out of equilibrium. In particular, the paper addresses the role of transaction innovations as well as the appearance and disappearance of transactors (new and old actors) induced by them.

Understanding long-run economic evolution as a step-wise evolution of money towards an ever increasing abstraction is the concern of Hardy Hanappi (Vienna University of Technology) in his contribution on “Money, Credit, Capital and the State: On the evolution of money and institutions”. This paper suggests that the development of money forms (exchange money, credit money, and capital) is closely linked to the development of social institutions, in particular, state institutions from feudal forms towards more continental political entities. Applying this co-evolution to the current general crisis indicates just a symptom of a mismatch between the available institutional systems and the capital program of large scale private firms.

In “Asymmetric markets and the evolution of the division of labor” Carl Christian von Weizsäcker (Max Planck Institute for Research on Collective Goods, Bonn) formulates an economic system that evolves according Darwinian variation-retention-selection dynamics. Individual freedom of choice is considered a constituting ingredient with respect to all three aspects of the Darwinian dynamics. Yet the economic system is characterized by a fundamental asymmetry: whereas the supply side of the market consists of competing firms, the demand side consists of individuals or firms that choose among the competing suppliers without being under competitive pressure themselves. This asymmetry is shown to be a necessary condition for an economy to be innovative.

The concluding paper in the volume shows that not economies evolve, but also our knowledge about evolutionary processes. In “Recent developments in evolutionary biology and their relevance for Evolutionary Economics” Karin Knottenbauer (RWTH Aachen) draws the attention towards recent findings in evolutionary biology that highlight the phenomena of co-operation, communication, and self-organization as sources of novelty. For the variation-retention-selection principle or mechanism, often at the core of evolutionary economic thinking, these developments suggest that selection is not the predominant factor of evolution, but only one among many. The implications for evolutionary economics with respect to analogies, generalized Darwinism, and the continuity hypothesis are also addressed.

Taken together, the papers in this volume depict a necessarily selective but hopefully also representative perspective on the “perennial gales of creative destruction” as envisaged by Schumpeter. Certainly, research along these lines has still been devoted to creativity issues of innovation processes and our insights have continuously advanced in this place. A more complete picture, however, should take the destructive aspects of an innovation-driven development on board and should emphasize the intertwinement with creative aspects of innovative activities. Since the literature concentrating on these topics is still rather sparse and in its infancy, we hope that the papers collected in this volume serve as a trigger for further research along these lines.

Mai 2013, Augsburg, Berlin, Jena and Kassel: Guido Buenstorf, Uwe Cantner, Horst Hanusch, Michael Hutter, Hans-Walter Lorenz, Fritz Rahmeyer.

References

Metcalfe S (2008) Restless capitalism—the evolutionary nature of competition. Princeton University Press, Princeton, NJ

Schumpeter JA (1994 [1942]) Capitalism, socialism and democracy. Routledge, London

Part I
Innovation: Conditions to Successfully Create Novelty for Economic Development

Agents of Change

Caroline Gerschlager

Abstract Evolutionary accounts in economics have offered a new look at economic development. Advancing on the limited cognitive capacities, they have shown that structural change follows technological, institutional and ideological paths. The present examination suggests an add-on to this cognitive portrayal of change. Drawing on Schumpeter's theory of innovation and Adam Smith's theory of knowledge, it analyzes human motives as important drivers of development. It brings in the dynamic nature of human motives and particularly discusses the human will as a requirement of change.

> "The difficulty lies, not in the new ideas, but in escaping from the old ones."
> (John Maynard Keynes: (2010[1936]) *The General Theory of Employment, Interest and Money*)
> "Der bloße Gedanke allein genügt nicht und setzt sich nie 'von selbst' durch... Der Vorgang ist vielmehr in der Regel der, dass der neue Gedanke von einer kraftvollen Persönlichkeit aufgegriffen und durch ihren Einfluss durchgesetzt wird.
>
> Jahrhundertelang kann eine neue Möglichkeit, trotzdem dass sie in recht weiten Kreisen bekannt ist, ein unfruchtbares Schattendasein führen, ohne irgendeine Wirkung nach außen zu haben."
> (Joseph Alois Schumpeter: 1911–12: *Theorie der wirtschaftlichen Entwicklung*)

Reprinted from Journal of Evolutionary Economics 22(3), 413–441, Springer (2012).
This version was presented at the 14th ESHET (European Society of the History of Economic Thought) Conference at Bosporus University in Istanbul, TR (19–21 May 2011) and the EAEPE Conference on *Schumpeter's Heritage* at the Techgate in Vienna, A (17–30 October 2011). A previous version was presented at Ausschuss für Evolutorische Ökonomie in Augsburg, G (10–12 July 2008). The author gratefully acknowledges comments from the members of the Ausschuss für Evolutorische Ökonomie and anonymous referees.

C. Gerschlager (✉)
Department Volkswirtschaft, Wirtschaftsuniversität Wien, Wien, Austria

DULBEA, Université Libre de Bruxelles, Bruxelles, Belgium
e-mail: caroline.gerschlager@wu.ac.at

G. Buenstorf et al. (eds.), *The Two Sides of Innovation*, Economic Complexity and Evolution, DOI 10.1007/978-3-319-01496-8_2,

1 Introduction

The economic agent in evolutionary economics is often characterized in contrast to the standard neoclassical view, and hence in contrast to rational choice (Vanberg 2002). In particular, evolutionary accounts stress fundamental uncertainty in decision making and consider imperfect information. Human action is defined in terms of rule following, and therefore is seen as the result of the past, rather than of the outcome of forward looking calculation of consequences; also, it does not impose optimization as a selection criterion.[1] Decision making and action are rule governed, and these rules are seen as the outcome of long run evolution of human action and individual learning.

With these assumptions, very important results about structural - economic and institutional - change have been gained. One of these is that economic development is path dependent: it follows technological (Nelson and Winter 1982; Arthur 1994), institutional (North 1999) and ideological paths (Denzau and North 1994; Denzau et al. 2007). Individual agents who are constrained by these paths can resist changing and contribute to inertia (and lock-in). Against this background, the manifestation of innovation is seen as a break from prevailing routines (Nelson and Sempat 2001: 44).

These analyses in general do not reveal much about the individual agent. This is because the focus of evolutionary analysis is meso and not micro (Dopfer 2006) and the attention is on the results of interaction among individuals within a population. The individual remains much of a black box, as Nelson and Winter (1982) make clear, since, for their purposes, it is not necessary to dig deeper into individual persons, that is, their character, and to know more about their preferences and beliefs (cf. also Vromen 2001).

Nelson and Winter (1982) emphasize that heterogeneous agents affect economic development and institutional change. They focus on competitive markets or other institutions to which the agent adapts or learns to adapt and in which bounded rationality in the sense of rule following has taken the place of perfect optimizing rationality as a selection mechanism. While approaches to path dependency assume heterogeneity of agents - for example, that some of them are more willing to change and break with the rules than others - the basis of their heterogeneity, including the manner in which their motives are formed and ways in which they might change, is typically not further explored. As a consequence, the human agent is moved to the background. And yet, it is finally the motivation of the agent that influences his decision to break eventually with the rules and routines and change.

Ideologies and institutions do not only shape individual agents and their preferences, as is typically recognized from an evolutionary and institutional perspective

[1] There is a shift of emphasis from an outcome oriented view, which relies on the rational prowess of agents and their deductive capacities, to a procedural view of the economy, in which the diversity of agents, their multiple expectations and beliefs as well as learning processes are the focus.

and against the standard view of rational choice in which stable agents with given preferences are assumed (Bowles 1998; Denzau and North 1994). Individual agents also shape and chose their ideologies and institutions. Accordingly, North (1999 [1990]) argues that the change of preferences played at least some role in the abolition of the economic institution of slavery. This view holds the promise that a better understanding of human motivation can also increase our understanding of change.

The present paper is motivated by this reverse causation. It attempts to dig a little deeper into the motivational foundations of change. To see clearer on this issue on which most economists today have little to say,[2] I will bring in Joseph Schumpeter and Adam Smith - two eminent economists of change and economic transformation -because of their strong views on the problem. Their historical accounts are seen here as treasure troves of ideas (Kurz 2006: 476) that can be used in "new combinations" (Schumpeter 1963[1934]: 66, 64) with a view to enriching current debates. I will by example draw on the entrepreneur of Schumpeter's *Theory of Economic Development* (1911–12, 1927, 1934) and the poor man's son of Smith's parable in *The Theory of Moral Sentiments* (2001[1759]).

One essential requirement of innovation and change is to break with the past - with previous thought and habits - and it is certainly the case that the human imagination is crucial in this regard.[3] This is so because, thanks to the imagination, the human agent can "entertain a world view that no longer relies exclusively on experience," as Dopfer puts it. In contrast to other primates, the human agent is free to "image a future that is disconnected from the past" (Dopfer 2004: 190). The human agent is capable of "originating, adopting and retaining novel rules" (Dopfer and Potts 2008: 30), and hence "of becoming generically different" (31). In this generic view, the human agent "creates, alters and abandons new rules" or knowledge (Grebel 2008: 7).

On the other hand, having this capability does not mean that the agent will use it when necessary. There is indeed nothing - from an evolutionary perspective at least - that guarantees that the agent will also carry out his unique capability.

My discussion of historical accounts suggests that knowledge in and of itself is not sufficient. What is also essential is the will. For an agent of change, therefore, it is not sufficient to know that, when and how to break with existing rules and routines, but he also must be *willing* to do so.

That the new is connected with the elimination of the old has been repeatedly affirmed: in his study on the principles leading to the advancement of

[2] Economists have long shown little interest in a better understanding of motivational foundations in general (Giocoli 2003; Klamer 1989). They usually apply a narrow concept of motives (Blume and Easley 2008; Loasby 2009a: 38).

[3] The recognition of the foundational role of the human imagination today is most prominently associated with the George Shackle (Shackle 1972; cf. also Loasby 1994). That change necessitates paying more attention to the imagination and going beyond rational choice has recently also been acknowledged by Denzau and North (1994) and March (1995). Cf. Augier and Kreiner (2000) and Patalano (2007) for further discussion.

understanding, Adam Smith emphasizes that the new system of Copernicus plainly "*destroyed* the system of Ptolemy" (1980[1795]): 84, italics by the author). Shackle likewise acknowledges that "new knowledge is in part destructive of old knowledge" (Shackle 2009[1970]: 21; cf. also Loasby 1994: 523) — since it is genuinely new, it "must subvert to some degree what has been accepted as knowledge hitherto" (Shackle 2009[1970]: 21). When maintaining in this context that the "vital aspect of the process of ... innovation is its destructive effect" (Shackle 1967: 295), Shackle almost strikes the Schumpeterian chord.

But yet, old knowledge or rules do not disappear effortlessly. Merely because something has become obsolete does not mean that it will also give way. The process of destruction generally creates resistance. The old and obsolete fights against losing its significance - against being eliminated, transformed or circumvented. For change and innovation to occur, therefore, it is essential that the resistance of the past is overcome. As Shackle makes clear in relation with theoretical innovation, the theoretician needs to "overturn the intellectual dwelling places of hundreds of people, whose first instinct will be revenge and resistance" (1967: 295).

For that reason, extra effort of the will is needed. The present examination of human motives can, therefore, also be seen as an inquiry into the significance of the human will and should be understood as complementary to work in which the stress is on processes of thought.[4] To be sure, it in no way neglects the significance of cognition. As has been argued above, the recognition of the limitations and capacities of human cognition is important to understanding economic behavior.

Hence change here is seen as willed. This implies that it is not the consequence of changing constraints as claimed by standard accounts. It is also not the result of error, history or chance. But change, at least to some degree, also involves an element of choice. One purpose of the paper, more generally, is to call attention to this 'some degree' which involves the will.

Schumpeter was fervent in defending the will. To set the scene, I will therefore discuss his theory of innovation. Adam Smith made the case for the advancement of human understanding. To emphasize the will's part in his theory of knowledge, I will particularly discuss the economic setting to which Smith connects it in *The Theory of Moral Sentiments* (2001[1759]).

The remainder of the paper proceeds as follows. Section 2.1 refers to Schumpeter's entrepreneur and discusses the resistances against the improvement of economic services. Section 2.2 refers to the poor man's son of Smith's parable and discusses the resistances against the advancement of understanding. Section 3 considers complementarities and contrasts between both theories. Section 4 concludes.

[4] Morrison and Potts (2008) refer to systematic cognitive biases or errors of human thought to explain why innovation processes are subject to failure. They suggest that these biases explain why it is very hard – and, in many cases, perhaps too hard one could advocate - for the agent to innovate. For a discussion and appraisal of the domain limitedness of knowledge and its evolutionary foundation in the human mind, cf. Loasby (2003: 33-5).

2 Willing to Change

2.1 Schumpeter's Entrepreneur and Economic Development

The breaking with routines is at the heart of Schumpeter's conception of agency. He develops this conception in contradistinction to rational behavior at the foundation of Walras' equilibrium theory of perfect competition. For Schumpeter, Walras' conception of rational behavior should be seen as a result of behavioral routines. The "assumption that conduct is . . . rational is in all cases a fiction. But it proves to be sufficiently near to reality, if things have time to hammer logic into men" (1963 [1926]: 80).[5]

Economic agents on this view act according to routines or rules that are, to a great extent, based on the experience of the past: "The individual household or firms acts, according to empirically given data and in an equally empirically determined manner. Obviously this does not mean that no change can take place in their economic activity. The data may change, and everyone will act accordingly as soon as it is noticed. But everyone will cling as tightly as possible to habitual economic methods and only submit to the pressure of circumstances as it becomes necessary. Thus the economic system will not change capriciously on its own initiative but will be at all times connected with the preceding state of affairs" (1963[1926]: 8-9).[6]

Similarly, in the first edition of *Theorie der wirtschaftlichen Entwicklung* (1911–12), Schumpeter refers to agents who are not interested in change as follows. "The data which have ruled the economy in the past are known, and if they remained stable, the economy would repeat itself in the same way. If the data change, this is not known in the same way, but the agent responds to them as well as he can. The agent does not change . . . anything on his own" (translation by the author).[7]

[5] "Annahme eines Verhaltens, das der Beobachter als prompt und rationell begreifen kann, ist eine Fiktion auf alle Fälle. Aber sie bewährt sich dann, wenn und weil die Dinge Zeit haben, Logik in die Menschen zu hämmern" (Schumpeter 1926: 118).

[6] "Das Wirtschaftssubjekt handelt also nach erfahrungsgemäß gegebenen Daten und in einer ebenso erfahrungsgemäß gegebenen Art und Weise. Natürlich heißt das nicht, daß keine Veränderungen in seiner Wirtschaft eintreten können. Die Daten derselben können sich ändern, und jedermann wird sich danach richten, sobald er es merkt. Aber dann wird jedermann nicht etwas schlechthin Neues tun, sondern möglichst viel von seiner gewohnten Wirtschaftsweise festhalten und dem Drucke der Verhältnisse nur soweit nachgeben, als nötig ist. Und auch dieses 'Nachgeben' wird er nach den Regeln der Erfahrung vollziehen. So würde sich das Bild der Wirtschaft nicht willkürlich ändern, sondern sich in jedem Augenblick an den vorgegebenen Zustand anschließen" (Schumpeter 1926: 7–8).

[7] "Die Daten, die die Wirtschaft in der Vergangenheit beherrscht haben sind bekannt, und wenn sie unverändert blieben, so würde die Wirtschaft in derselben Weise wieder ablaufen. Die Veränderungen, die sie erleiden mögen, sind nicht ganz so bekannt, im Prinzipe folgt ihnen das Wirtschaftsubjekt so gut es kann. Es ändert . . . nichts selbsttätig" (1911–12: 32).

According to Schumpeter, standard economics describes the manner in which economic agents react to given constraints, and he shows that these reactions are predetermined. In a static economy, which is based on what we today call rational choice theory, nothing new can happen, since agents can only passively adapt to the data. Even if the data change, nothing new can arise, because the agents ("Wirte" in Schumpeter's terminology) optimally adapt to these changes. They satisfy their wants given the constraints they face. They are in a circular flow, Schumpeter contends, populating an economic environment that is "resting" and "passive", a "static, stationary economy determined by the circumstances (1911–12: 87) (translation by the author).[8] These attributes are metaphorical because the circle does not exclude accumulation and growth. But Schumpeter emphasizes that economic growth as a result of growing wealth or population will not result in a change of routines. It will primarily result in behavior that adapts to the change in data. Such an economy is populated by people who always remain the same, and of goods that always are reproduced in the same way (Gerschlager 1996: 117-8).

From this follows the notion that innovations, if they appear at all in the economic system, are disruptive of these behavioral routines, and hence the economic system. The disruptive, disturbing, and finally always destructive nature of the new figures prominently in his theory of the business cycle (1934[1963]: 216, 252).

Schumpeter draws attention to the evolving character of the economy. He seeks to explain the structural changes that are observed in economic history and interprets these as movements away from equilibrium. He is interested in the shifting of equilibriums and in discontinuity (1926: 99).

For Schumpeter, the challenge, therefore, is not to "combine materials and forces within our reach… but to produce other things, or the same things by a different method, means to combine these materials and forces differently" (1963 [1934]: 65).[9] Form and content of development is defined in terms of "carrying out of new combinations" (66).[10] This involves new goods, new methods of production, new markets, new resources, new organizational structure, i.e. the creation or breaking up of a monopoly position (66).

Against this view of the agent, he brings in the entrepreneur as the agent who realizes the new combinations and triggers change. He is a not a dreamer but a "Mann der Tat" (1911–12: 32) - a man proceeding to action. He has to swim "against the stream if he wishes to change its channel" (79-80),[11] and what "was formerly a help becomes a hindrance. What was a familiar datum becomes an unknown. Where the boundaries of routine stop, many people can go no further, and

[8] "ruhend, passive, von den Umständen bedingte, stationäre, statische Wirtschaft" (1911–12: 87).

[9] "vorhandene Dinge und Kräfte kombinieren" …"anderes oder anders produzieren heißt diese Dinge und Kräfte anders kombinieren" (1926: 100).

[10] "Durchsetzung neuer Kombinationen" (1926: 100).

[11] "gegen den Strom, wenn er dessen Bahn verändern will" (1926: 118).

the rest can only do so in a highly variable manner" (80).[12] According to Schumpeter, leadership exists because "every step outside the boundary of routine has difficulties and involves a new element" (84).[13]

The difficulties arise from limited information and uncertainty, particularly in dealing with the new data. "Now he must really to some extent do what tradition does for him in everyday life, viz. consciously plan his conduct in every particular. There will be much more conscious rationality in this than in customary action, which as such does not need to be reflected upon at all; but this plan must necessarily be open not only to errors greater in degree, but also to other kinds of errors than those occurring in customary action. What has been done already has the sharp-edged reality of all the things which we have seen and experienced; the new is only a figment of our imagination. Carrying out a new plan and acting according to a customary one are things as different as making a road and walking along it" (85).[14]

Another difficulty is related to the fact that the agent of change has to overthrow two principal forms of resistance. First, he has to fight against the reluctance to change that arises from within his own person: "In the breast of one who wishes to do something new, the forces of habit rise up and bear witness against the embryonic project. A new and another kind of effort of will is therefore necessary in order to wrest, amidst the work and care of the daily round, scope and time for conceiving and working out the new combination and to bring oneself to look upon it as a real possibility and not merely as a day-dream. This mental freedom presupposes a great surplus force over the everyday demand and is something peculiar and by nature rare" (86).[15] Second, he has to overcome the resistance from the social

[12] "dort Stütze war, wird hier Hindernis. Was vertrautes Datum war, zu einer Unbekannten. Wo die Grenze der Routine aufhört, können deshalb viele Leute nicht weiter und der Rest kann es nur in sehr verschiedenem Maß" (1926: 118).

[13] "jeder Schritt aus dem Bezirk der Routine Schwierigkeiten hat, ein neues Moment involviert" (1926: 124).

[14] "Wohl handelt er auch nach einem Plan; es wird sogar viel mehr bewußte Rationalität darin stecken als im gewohnten, der als solcher überhaupt nicht 'überlegt' zu sein braucht; aber dieser Plan muß erst erarbeitet werden. Und deshalb enthält er nicht bloß graduell größere, sondern auch andere Fehlerquellen als der gewohnte. Der gewohnte hat die ganze scharfrandige Realität der Vorstellungen von Dingen, die wir gesehen und durchlebt haben; der neue ist eine Vorstellung von Vorgestelltem. Nach ihm handeln und nach dem gewohnten handeln sind so verschiedene Dinge wie einen Weg bauen und einen Weg gehen: Und das Bauen eines Weges ist so wenig ein bloßes gesteigertes Gehen, als das Durchsetzen neuer Kombinationen ein bloß graduell vom Wiederholen der gewohnten verschiedener Prozeß ist" (1926: 124-5).

[15] "In der eigenen Brust dessen, der Neues tun will, erheben sich die Elemente der gewohnten Bahn und legen Zeugenschaft ab gegen den werdenden Plan. Eine neue und andersgeartete Willensaufwendung wird dadurch nötig, außer jener, die schon darin liegt, inmitten der Arbeit und Sorge des Alltags um Raum und Zeit für Konzeptionen und Ausarbeitung der neuen Kombination zu ringen und sich dahin zu bringen in ihr eine reale Möglichkeit und nicht bloß Traum oder Spielerei zu sehen. Diese geistige Freiheit setzt einen großen Überschuß von Kraft über das Erfordernis des Alltags voraus, ist etwas Einzigartiges und ihrer Natur nach selten" (1926: 126).

environment. It is a feature of every social environment to have a general tendency to condemn “any deviating conduct by a member of a social group” (86).[16] The entrepreneur has to find “the necessary cooperation” (87)[17] to overcome opposition particularly from those “groups threatened by the innovation” and from “the difficulty in winning over consumers” (87).[18]

The function of leadership does not lie in “’find[ing]’” or “’creat[ing]’ new possibilities.” These new possibilities, according to Schumpeter “are always present, abundantly accumulated by all sorts of people”; for the most part “they are also generally known and being discussed by scientific or literary writers” (88) - but the leader’s function is to “do… the thing” (88).[19] Schumpeter stresses that leadership concerns the bearing that change has on others. In other words, others have to follow and change, too. “The type of the leader is characterized by his specific way to look at things - and thereby not his intellect is important but his will, his energy, to tackle certain things and to see them as [if they were] real -, his ability to go on his own and to go on ahead, not feeling uncertainty and resistance as counterarguments, and finally the effect he has on others” (translation by the author).[20]

Schumpeter’s thick description of the entrepreneur has to be understood in the light of his attempt at formulating the laws of action within the larger framework of a theory of action (1911–12: 134).[21] He maintains that, in relation with change, the economic agent is to be conceived differently than in a static economy. Gossen’s laws and the rationality of the satisfaction of wants they imply - that the intensity of the agent’s wants would decrease with increasing quantity of the good consumed for example - are only pertinent in a static context but are secondary in a changing economy. In the latter, economic behavior is to be seen as depending on the evolution of the agents’ wants.

Change can be better understood, according to Schumpeter, if we attribute greater significance to individual motives. This is in contrast to the analysis of the

[16] “mißbilligt” … “jedes abweichende Verhalten eines Gliedes der sozialen Gemeinschaft” (1926: 127).

[17] “erforderliche Kooperationen seitens der Leute zu finden, die man braucht” (1926: 127).

[18] “durch das Neue bedrohten Gruppen” … “ in der Schwierigkeit die Konsumenten zum Mitgehen zu bringen” (1926: 127).

[19] “Der Führer als solcher “findet” oder “schafft” die neuen Möglichkeiten nicht: Die sind immer vorhanden, reichlich angehäuft von Leuten im Lauf ihrer gewöhnlichen Berufsarbeit, oft auch weithin gekannt und, wo es Literaten gibt, auch propagiert.” … “Die Führerfunktion besteht darin, sie lebendig, real zu machen, *durchzusetzen*” (1926: 128, italics in original).

[20] “[D]er Typus des Führers ist charakterisiert einmal durch eine besondere Art die Dinge zu sehen – dabei wiederum nicht so sehr durch Intellekt (und, soweit durch diesen, nicht einfach durch Weite oder Höhe, sondern gerade durch eine Enge bestimmter Art) als durch Willen, durch die Kraft, ganz bestimmte Dinge anzufassen und sie real zu sehen –, durch die Fähigkeit, allein und voraus zu gehen, Unsicherheit und Widerstand nicht als Gegengründe zu empfinden, und sodann durch seine Wirkung auf andre” (1926: 129).

[21] “aus den Motiven ein Verständnis und mit ihm ein Gesetz des Handelns zu gewinnen“ (1911–12: 134).

circular flow, "where the importance of examining motives is very much reduced," since "the equations of the system of equilibrium may be so interpreted as not to imply any psychic magnitudes at all, as shown by the analysis of Pareto and of Barone" (1963: 90-1). Motives also characterize the behavior in question. In this light, Schumpeter calls the entrepreneur "most egotistical of all" (1963: 91) and "reckless" (translation by the author).[22] He "is more self-centred that other types because he relies less than they do on tradition and connection" (91-2).[23]

When making the decision to carry out new combinations, "he must consciously plan his conduct in every particular" (1963: 85), because he can no longer rely on knowledge and rules of conduct within the accustomed channels. In Schumpeter's perspective, the entrepreneur, therefore, is not only the most egotistical but also the "most rational of all," since to carry out "new plans, requires more conscious rationality than the mere running of established business... which is largely a matter of routine" (92). The entrepreneur as Schumpeter sees him will not act, unless he also can "seize a gain" (214). A "tolerably reliable calculation" of "new possibilities more advantageous from the private economic standpoint" (214) is therefore imperative. Schumpeter is well aware that calculation necessitates economic stability. In times of disturbance - when economic development comes to a halt - the entrepreneur cannot obtain the security he needs to know that it is worthwhile to move ahead.

That the new should, therefore, not be seen in direct contrast to the old as Schumpeter claims, but as depending on it, has been repeatedly urged (Festré and Lazaric 2004). Neither the calculation of an advantage of new combinations (Loasby 2003: 176) nor their imagination is possible without "the background of relevant stability"; for "change cannot [even] be recognized" without it (Loasby 2005: 14).

Indeed, the stability of (old) rules or knowledge is significant; it is also an important precondition for what most interests me here in relation with entrepreneurial action: the human will. For, if rules did not importantly exist, the will to break with them would be pointless.[24]

We might, therefore, wonder whether Schumpeter has given "adequate consideration to the elements of continuity which are necessary to carry even radical transformation" (Loasby 2005: 14)? In this light we should also be inclined to ask whether Schumpeter's depiction of his protagonist was not "too dramatic," as

[22] He refers to the entrepreneur as an egoist, in the sense of "gesteigertem Egoismus" and "Rücksichtslosigkeit" (1926: 134).

[23] "His characteristic task ... consists precisely in breaking up old" ... "tradition" and although "this applies primarily to his economic action, it also extends to the moral, cultural, and social consequences of it" (1963: 92).

[24] When making the case for the precedence of norms over rational choice Elster applies a similar argument: if people did not believe in the existence of norms, he says, there would not be anything for them to manipulate (1989: 34-5).

Penrose puts it (quoted by Loasby 2003: 180). I think that, besides effective rhetoric (Loasby 2003: 177), Schumpeter's insistence on contrasts and "the drama" (1911–12: 502) of change he vividly and repeatedly portrays are well founded after all. Since they are based, I believe, on the realization that destruction is quite real in change, the resistance it breeds needs to be overthrown. From the point of view of the innovator as a destructor, the new truly *is* in sharp opposition to the old. After all, his job is to destroy and hence to bury old hopes and values forever (Schumpeter 1963: 217).[25]

Schumpeter therefore insisted on the special effort of the will. For "doing the thing" for "setting up new production functions," as Schumpeter also put it in his *Business Cycles*, the will is needed to overthrow those "who . . . will fight hard to preserve their culture and status" (McCraw quotes Schumpeter 2006: 242).

This ambivalence of change leads to a fairly disgraceful description of its protagonist. The "entrepreneurial kind of leadership" . . . "has none of that glamour which characterises other kinds of leadership", Schumpeter assesses, and it "consists in fulfilling a very special task which only in rare cases appeals to the imagination of the public. For its success, keenness and vigor are not more essential than a certain narrowness which seizes the immediate chance and *nothing else*. . . . He renders a service... not so easily understood by the public at large as a politician's successful speech or a general's victory in the field, not to insist on the fact that he seems to act - and often harshly - in this individual interest alone" (1934[1963]: 89, italics in the original).

Therefore, insisting that the job of destruction is to be done is different from glorifying it. Schumpeter makes clear that he does not intend to "style every entrepreneur a genius or a benefactor to humanity nor . . . to express any opinion about the comparative merits of the social organisation in which he plays his role, or about the question whether what he does could not be effected more cheaply or efficiently in other ways" (1934[1963]: 90fn1).[26]

The emphasis on contrasts and oppositions notwithstanding, stability and instability, i.e. order and disorder, depend on each other, and I think this is the case in Schumpeter's account, too. Schumpeter further develops this in his theory of the business cycle, with reference to the necessary interplay between periods of boom and depression. On this view, situations of stability or boom are succeeded by situations of economic instability or depression in which subjects are constrained to adapt to the new standards the boom introduced. Not until these and hence stability and a new equilibrium are reached (1963: 244-5) can the entrepreneur "see" future

[25] The "economic system needs rallying before it can go forward again. Its value system needs reorganizing. And the development which then starts again is a new one, not simply the continuation of the old. It is true, experience teaches that it will move more or less in a similar direction to the earlier, but the continuity of the 'plan' is interrupted. The new development proceeds from different conditions and in part from the action of different people; many old hopes and values are buried forever, wholly new ones arise" (1963: 217).

[26] His aim was to distinguish his agent of change from that of a "robber" for example (1934[1963]: 90fn1).

profits. In the meantime, the impossibility of calculation is enough to inhibit entrepreneurial activity, even of the strongest willed.

For Schumpeter, the destruction of old values and the ensuing economic instability is unavoidable, since the "new requires that those who wander along their well known paths change their ideas" (translation by the author).[27] This is why "neither profits in a boom nor losses in a depression are meaningless and functionless" (252-3). "[T]hey are essential elements of the mechanism of economic development and cannot be eliminated without crippling the latter" (253). Schumpeter's analysis of the business cycle is generally seen as important argument against Keynesian claims for intervention.

In contrast to Keynes, however, Schumpeter was reluctant to give unmistakable policy advice. In his view, his analysis can "be used to derive practical conclusions of the most conservative as well as the most radical complexion." As McCraw (2006) reminds us, Schumpeter is eager to point out that it did not lend support "to any general principle of *laisser faire*" (with reference to Schumpeter's *Business Cycles*, vol. 1, vi: 237 f. 14, italics in original).

Moreover, despite the fact that Schumpeter insisted on destruction as part of economic development, he does not tell how important the destruction has to be. In the final pages of his *Theory of Economic Development*, he emphasizes that losses and destruction can also be "meaningless and functionless" (1963: 253). By so doing, he also leaves potential room for new forms of intervention to mitigate if not to avoid some of the destruction deemed useless. I will, however, not further discuss policy perspectives here as Schumpeter was not importantly concerned with them.

Let us therefore disregard the question whether the process of destruction has to be completed entirely or in degree; but let us assume that order and stability set in with or without intervention. Let us further assume that the general conditions for innovation are met. The entrepreneur can see the gain his idea will bring "immediately before his eyes" (214).

Even now, this is not sufficient. Schumpeter emphasizes that the entrepreneur must also "break up" (92), i.e., he must be willing to destroy; for if not, the new will not find approval, as "it would be objected or it would only receive such weak and vague acceptance that can never be successful and fertile" (translation by the author).[28]

According to Schumpeter, the motivation of the entrepreneur is different from that of most other agents (1926: 134). "The typical entrepreneur does not ask whether his effort is gratified. He does not care about the hedonic satisfaction that can be derived from his deeds. He works restlessly because he cannot do otherwise, he does not live for enjoying what he has acquired. Should this wish arise, then it is

[27] Für die Aufnahme von etwas Neuen ist ein Prozess des Umdenkens für alle in statischen Bahnen Hingleitenden nötig (1911–12: 543-4).

[28] "er würde auf Ablehnung oder doch nur auf jene matte, vage Art der Zustimmung stoßen, die zu wirklicher Fruchtbarkeit nie führen kann" (1911–12: 543-4)

an indication of an end and not an interim target, heralding physical death rather than satisfaction" (translation by the author).[29]

On Schumpeter's view, the type of the entrepreneur is best to be identified by his "Motto," which is "plus ultra" (1926: 137). Plus ultra is first motivated by "the dream and the will to found a private kingdom . . . a kingdom that provides for the sensation of power and independence . . . and which is specially strong for people who have no other chance of achieving social distinction" (1963[1926]: 93).[30] Second, Schumpeter refers to "the will to conquer: the impulse to fight, to prove oneself superior to others, to succeed for the sake, not of the fruits of success, but of success itself."[31] Exemplary for this motivation is to look at economic action in terms of sports: "there are financial races, or rather boxing-matches" (1963[1926]: 93) also involving "social ambition" relating to the first motivation.[32] Finally, Schumpeter emphasizes a third family of motives: "the joy of creating, of getting things done, or simply of exercising one's energy and ingenuity" (93). For the entrepreneur, change is seen as an art pour art. He "seeks out difficulties, changes in order to change, delights in ventures" (93-4).[33]

All of these motives are active and this is what, according to Schumpeter, distinguishes them from the standard view in which the agent is primarily seen as reactive, i.e. reacting to internal or external constraints. The entrepreneur does not adapt to the environment, but attempts to change it in the first place. This change is to be understood in relation with these motives.

Since the theory of concrete interest fails in Schumpeter's view, he explores whence the "enduring disposition [of the entrepreneur] seizing upon one opportunity as eagerly as the next" (1966[1919]: 6). He shows that this disposition is driven by some instinctive "urge to action." The entrepreneur's "urge to domination" (12), for example, springs from the "capacities and inclinations that had once been crucial to survival" (33; cf. also Festré and Lazaric 2004: 19-20). The entrepreneur is "never satisfied by the fulfillment of a concrete interest, as would be the case if fulfillment were the motive... ... Hence the tendency of such expansion to transcend all bounds and tangible limits, to the point of utter exhaustion." In his

[29] "Der typische Unternehmer frägt sich nicht, ob jede Anstrengung, der er sich unterzieht, auch einen ausreichenden 'Genußüberschuß' verspricht. Wenig kümmert er sich um hedonische Früchte seiner Taten. Er schafft rastlos, weil er nicht anders kann, er lebt nicht dazu, um sich des Erworbenen genießend zu erfreuen. Tritt dieser Wunsch auf, so ist das Erlahmen und nicht eine Station auf bisheriger Linie, Vorbote des physischen Todes und nicht Erfüllung" (1926: 137).

[30] "Traum und Wille, ein privates Reich zu gründen... ein Reich, das Raum gewährt und Machtgefühl... und dessen Faszination gerade für solche Leute besonders wirksam ist, die keinen Weg zu sozialer Geltung haben" (1926: 138).

[31] "Siegerwille, Kämpfenwollen einerseits, Erfolghabenwollen des Erfolgs als solchen wegen andererseits" (1926: 138).

[32] "Finanzieller Wettlauf, noch mehr aber Boxkampf." "[S]ozial Steigenwollen" (1926: 138).

[33] He changes the economy "um des Änderns und Wagens und gerade der Schwierigkeiten willen" and because of "Freude am Werk, an der Neuschöpfung als solcher: Sei das nun etwas Selbständiges oder ununterscheidbar von der Freude am Tun" (1926: 139).

"objectless disposition to expand ... without limits"... the modern captain of industry for Schumpeter resembles the warrior and hunter" (1966[1919]): 29 [fn 1: 171]). The predominance of instincts in the mentality and the "unrestrained will to gratify these instincts" (33) are seen as remnants of previous forms of social organization, the "warrior nation[s]" of the past (1966[1919]: 37).

Hence, the Schumpeterian agent of change is unconscious and driven by instinct rather than rational deliberation. For him, "instincts and interests" are the same (24). Schumpeter well understands that "human motivation" is "infinitely complex, and we are never aware of all its elements" (32).

Ideas are secondary in Schumpeter's account; first comes the agent and his will. Schumpeter is not really interested in knowledge (Pavitt 1998; Dopfer 2006) and has been criticized for not telling whence the entrepreneurial vision. However, considering that he was convinced that new knowledge existed in surplus (Schumpeter 1963[1934]: 88) it seems only natural that he did not give priority to explaining it. While Schumpeter is aware that entrepreneurship is not always successful (1934[1963]: 85, 222) - he acknowledges that "error must play a special role ... [in economic development]" - he is not really interested in failure, another lacuna for which his theory has been critically assessed.

Yet, Schumpeter makes clear why in his analysis "no 'error theory' will be found." (1934[1963]: 227). To explain the "blunders and destruction" (252) of the business cycle, error in his view "is indeed a supporting and accentuating circumstance, but not a primary cause necessary to the understanding of the principle. There would still be cyclical movements - though in a milder form - even if no one ever did anything that could be described as 'false' from his point of view; even if there were no technical or commercial 'error,' or 'speculative fever,' or groundless optimism and pessimism; and even if everyone were gifted with wide foresight. The objective situation which the boom necessarily creates explains exclusively the nature of the thing" (227-8).

Schumpeter is clearly conscious of the proximity between success and failure in a wider context, of which the term "creative destruction" he coined also bears testimony (1942[1975]: 81-6). It is true, though, that Schumpeter lacks a theory of knowledge generation. He confers little space to (failure as a consequence of) false knowledge and does not especially appreciate the positive conception of error key to the evolutionary understanding, to wit, making 'good mistakes' and making efforts to 'fail in a better way.'

Just as because Schumpeter does not proffer a theory of knowledge, Adam Smith has been criticized, not the least by Schumpeter (1954), because of a lack of a theory of innovation, that is, the absence of an explanation as to how new knowledge is effectively carried out.[34] But maybe Smith is of the opinion that "undertakers" and those people "who live by profit" (1997[1776]: 66, 265) were around in abundance anyway, and an account of entrepreneurial action is not considered as urgent.

[34] For a very insightful critique cf. Pesciarelli (1989).

On the other hand, and in contrast to Schumpeter, Smith was puzzled by the growth of knowledge. This was also appreciated by Schumpeter who is also more reconciling with his insights into human understanding, describing his early *Essay on Astronomy* (Smith 1980[1795]) as a "pearl" (Schumpeter 1954).

This *Essay*, which will figure as the point of departure of the next section, bears testimony of Smith's interest in the principles that lead man to advance his knowledge and ideas. In evolutionary accounts of economics, Smith is regularly used to fill the 'knowledge gap' Schumpeter left (Dopfer 2006; Pavitt 1998).

As we will see, for Smith, the difference between truth and falsity is important. In contrast to Schumpeter's theory of innovation, Smith's theory could even be seen as building on the significance of errors, i.e. false knowledge.

This also excludes the Humean skepticism about the possibility of establishing universal truths. Smith surely believed in the possibility of improvement, including the improvement of knowledge given appropriate motivation which is part of the human potential, and conditions.[35]

So, how can the inquiry into human motives increase our understanding of knowledge? And if so, how can this advance the present case made for the will?

2.2 Adam Smith's Poor Man's Son and the Development of Ideas

Long before Smith inquired into *The Wealth of Nations* (1997[1776]) and the economic utilization of knowledge, he elucidated its mental underpinnings. In the 1750*s*, he wrote an *Essay* on the principles that lead and direct "the natural progress of the mind in the investigation of truth" illustrated by the*History of Astronomy* (1980[1795]).

For Smith, reason and reflection are not enough to account for the occurrence of the new. His theory of knowledge growth is based on the imagination. What's more, it elaborates on the mental stimulus compelling the agent to use his imagination, when necessary.

According to Smith, the mind is well when it continues to think as it has always thought, for this is comforting. As Smith puts it, the imagination is "indolent" (1980 [1795]: 86). The mind, therefore, has a strong motivation to *not* innovate and change. To get going, a stimulus is needed. Smith makes clear that the agent will only use his mind if he perceives some discontinuity - "a gap, or interval" with the past. The ensuing "difficulty" (41-2), or discontent works as a stimulus for the mind to search for (new) ideas that re-establish continuity.

Smith depicts the contradictory motivations to innovate and not to innovate as complementary. Since the change can only be noticed against some continuity that is provided by the past. As noted above, this complementarity is also apparent in

[35] I owe the specification of this paragraph to the anonymous referee.

Schumpeter, whose entrepreneur needs the confidence that others will continue to rely on the past, in order to break with it. With his system of mental stimulus and response, Smith defends a dynamic account of knowledge founded on the limitations and capacities of the human mind. In this view, the human mind cannot help creating; it cannot help forever imagining new knowledge. Thus, ideas are abundant.

However, this is only one side of the coin. The other side is that the development of ideas can be importantly distorted or hindered. Smith is aware of "prejudices" (1980[1795]): 76) such as authority (67) and the reference of the past (Raphael and Skinner (1980[1795]: 14), which is why the French took so long to discard the Cartesian system, although Newton's system was closer to the truth; the "education" (76) and "the imagination" (86) or the "sense" (77) which sustained the belief that the earth is motionless, for example. As Raphael and Skinner note, these prejudices are indicative of "a certain unwillingness" and "even resistance" (1980 [1795]: 14). They hinder the advancement of understanding and, as a consequence, false knowledge abounds.

In *The Lectures on Rhetoric* (1983[1762]), Smith identifies a further prejudice: the "Newtonian method" according to which we "may lay down certain principles known or proved in the beginning, from whence we account for the severall Phenomena, connecting all together by the same Chain" (22). This superior method, first attempted by Descartes, particularly appeals to the human mind, giving "us a pleasure" (22).

At the same time, Smith warns that the apparent attraction of this method does not guarantee that the insights it supplies are without flaw and closer to the truth than the those supplied by other - perhaps less ingenious - methods such as the Aristotelian "unconnected method [Smith discusses in comparison], where everything is accounted for by itself without any referen[e]ce to the others" (136). The problem he identifies is that "the Cartesian Philosophy . . . tho it does not perhaps contain a word of truth . . . . . . nevertheless [has] been so universally received by all the Learned in Europe at that time" (136).

However, Smith was aware that advancement of understanding depends on the elimination of false knowledge. Otherwise, progress is impossible, as knowledge, although abundant, will not advance but deteriorate.

Smith's inquiry into the mental foundations of knowledge importantly analyzes the hindrances that are in the mind's way to understanding the world correctly. I will now turn to *The Theory of Moral Sentiments* (2001[1759]) because this allows me to discuss a singular hindrance to the growth of knowledge which as yet has not been mentioned: strong human motives. I bring in Smith's *Theory* because it is the treatise in which Smith most profoundly examines human motivation. I should mention here that this is also the treatise in which Smith concentrates on the way in which motives affect *moral* judgment, but here I am more interested in the effects they can have on human thinking, and the generation of knowledge in general.

Adam Smith is routinely seen as the originator of the economic agent who incessantly pursues his self-interest. Although this description of the agent is an inadequate simplification of the variety of human motives Smith acknowledges (Skinner 1996: 70; Rothschild and Sen 2006), it is clear that Smith also attributed much attention to the self-interested motives. I particularly concentrate on these to emphasize the difficulties they can bring about for the growth of knowledge. To specify what Smith had in mind when he referred to self-interest, I would like to recall that he uses the term interchangeably with self-love[36] the former denoting a desire for wealth, the latter for social approval and position. Most people are first and foremost concerned with themselves; in other words, according to Smith they seek approval, they want to be well established, and they want to be well-off; for Smith, all of this is only natural, since man is also "recommended by Nature" to his "care and attention" (Smith 2001[1759]: 219].

On the other hand, this concern can turn into an important difficulty, if it gets out of control. Vanity, avarice and ambition are cases in point. These strong motives are not indications of self-love or self-interest tout court, but of man's tendency to excess. These "extravagant passions" (2001[1759]: 149), as Smith also calls them, can deceive their subjects. Unlimited pursuit of wealth is indicative of avarice and insatiability regarding status and approval indicate ambition and vanity. The mind under the influence of these passions makes erroneous evaluations of outcomes of alternative actions. "Avarice overrates the difference between poverty and riches; ambition, that between a private and public station; vain-glory, that between obscurity and extensive reputation" (2001[1759]: 149).

In particular, the overestimation of approval, position and wealth results in the erroneous belief that these 'goods' exist without any bounds. Typically, ambitious and vain men are ignorant regarding the limits of position and approval. Besides, Smith also observes that their "misfortunes" arise "from their not knowing when they were well". According to him "the distress of disappointed avarice and ambition" is best summed up in "the inscription upon the tomb–stone of the man who had endeavoured to mend a tolerable constitution by taking physic; . . . '*I was well, I wished to be better; here I am*'" (150, italics in the original).

The concentration on human motives has allowed us to dig deeper regarding the origin of false beliefs. Accordingly, systematic errors of perception regarding wealth, approval and position find their origin in excessive self-love or self-interest. This is also why Smith refers to them as "delusions of self-love" (159). These false beliefs are primarily the result of the agent's character, and not of his mind.

I believe that such concern for human motives is at the heart of what I suggest calling Smith's 'deception theory'. It highlights strong motives and their influence on the human mind, and so further specifies the forces that are opposed to

[36] To wit, his famous statement: "It is not from the benevolence of the butcher, the brewer, or the baker, that we expect our dinner, but from their regard to their own interest. We address ourselves, not to their humanity but to their *self–love*, and never talk to them of our own necessities but of their advantages" (1997[1776]: 26-7, italics by the author).

knowledge growth. Smith's deception theory of *The Theory of Moral Sentiments* could therefore be seen as a special case of his theory of knowledge.

Smith develops his theory of mental stimulus and response with reference to scientific progress. He exemplifies the historical progress of our understanding of the movement of the celestial bodies and the earth by reviewing the merit and demerit of alternative theories which existed in the past. His theory, therefore, chiefly - but not exclusively - was conceived for the man of science or the philosopher as he is "particularly subject to ... [it], partly as a result of superior powers of observation and partly because of that degree of curiosity which normally leads him to examine problems ... which are to the ordinary man so 'familiar' as not to require any explanation at all." Smith considered that his mental account of making new connections was "typical of all men" (Campbell and Skinner 1997 [1976]: 2).

In view of that, I will discuss the poor man's son as an illustration of Smith's deception theory. The poor man's son seems an "ordinary man," as he does not aspire to wisdom but primarily is attracted to approval and position. He is not a learned man either, nor a scientist or philosopher. And yet, he attempts to better understand his situation and starts asking questions. Asking questions, Dopfer also asserts from the knowledge perspective distinguishes man from primates (2004: 191). I would think that this also involves the ability to ask questions of oneself.

But let us finally take a look at the story:

> "The poor man's son, whom heaven in its anger has *visited with ambition*, when he begins to look around him, admires the condition of the rich. ... He is displeased with being obliged to walk afoot, or to endure the fatigue of riding on horseback. He sees his superiors carried about in machines, and imagines that in one of these he could travel with less inconveniency. ... He thinks if he had attained all these, he would sit still contentedly, and be quiet, enjoying himself in the thought of the happiness and tranquillity of his situation. He is enchanted with the distant idea of felicity. It appears in his fancy like the life of some superior rank of beings, and, in order to arrive at it, he devotes himself for ever to the pursuit of wealth and greatness. To obtain the conveniences which these afford... He studies to distinguish himself in some laborious profession. With the most unrelenting industry he labours night and day to acquire talents superior to all his competitors. He endeavours next to bring those talents into public view, and with equal assiduity solicits every opportunity of employment. For this purpose he makes his court of all mankind; he serves those whom he hates, and is obsequious to those whom he despises... through the whole of his life he pursues the idea of a certain artificial and elegant repose which he may never arrive at, ... if in the extremity of old age he should at last attain to it ... it is then, in the last dregs of life, his body wasted with toil and diseases, his mind galled and ruffled by the memory of a thousand injuries and disappointments ... that he begins at last to find that wealth and greatness are mere trinkets of frivolous utility, no more adapted for procuring ease of body or tranquillity of mind, than the tweezer-cases of the lover of toys. ... In his heart *he curses ambition*, and vainly regrets the ease and the indolence of youth, pleasures which are fled for ever, and which he has foolishly sacrificed for what, when he has got it, can afford him no real satisfaction. In this miserable aspect does greatness appear to every man ... [who] observes with attention his own situation and ... consider[s] what is really wanting for his happiness" (Smith 2001[1759]: 181-2), italics by the author).

The protagonist of the parable clearly belongs to the breed of ambitious and vain men of the economy of "wealth and greatness" (2001[1759]: 181). Smith uses the story of the poor man's son in an important chapter of *The Theory of Moral Sentiments* directly related to economic development and the progress of humankind. I will come back to Smith's use of the parable at the end of this Section.

The poor man's son works hard. He succeeds in realizing his ambitions. Not only has he become rich, but he has also acquired status and social approval, and this, as he expected in his youth, has made his life more comfortable. Hence, we would expect that he "sit[s] contentedly." But despite social and economic success, he doesn't.

The puzzle of discontent of Smith's protagonist first and foremost interests me from the perspective of knowledge.[37] Smith tells that the poor man's son is disturbed, since his beliefs regarding position, approval and wealth have become obsolete, hence the discontent of his mind.

This cognitive discontent manifests itself in the realization of the false belief. The poor man's son becomes aware that his beliefs about wealth, approval and position no longer fit; he is mistaken to believe that approval and wealth are without bounds. His problem does not lie in not knowing what would be able to satisfy his ambitions, but in not knowing when these are satisfied. As Smith says, he did not know *when* he was "well, when it was … proper for him … to sit still" (2001 [1759]: 150, italics by the author).

We can dig deeper and inquire into the causes of his deception. From the perspective of Smith's deception theory, his false belief does not originate in biased cognition but in a biased character. The absence of limits is a difficulty that is essentially connected with ambition and vanity, for these extravagant passions are insatiable by nature. They are never satisfied, but grow with their stimulation. The parable, therefore, confronts its protagonist with his most profound inclinations. He understands that the passions that moved him ("he devotes himself for - ever") can also deceive him.

According to Smith's account of knowledge, the discontent of the poor man's son - the "gap" he presently experiences to his previous thought and action - should stimulate the mind to generate and to adopt new and more adequate ideas. But what precisely are these new ideas able to "soothe his imagination"? (Smith 1980[1795]: 46)

We do not know, since, we are not told what our protagonist does next. The story leaves us wanting with regard to his future projects. However, while incomplete with regard to the adoption and retention of new combinations, the story already makes a strong point with regard to the focus that is adopted here. It shows that false beliefs can hinder the agent in advancing his understanding. The story is instructive with regard to the abandonment of the past as determinant of change, and it also shows that abandoning obsolete ideas is not straightforward. Just because the agent knows that his beliefs have become obsolete this does not mean that he will also get

[37] That the parable is to be seen in the context of Smith's theory of knowledge has been pointed out by Loasby (2009b: 8-9).

rid of them. Just because the agent feels that his ideas are inadequate does not guarantee that he will lose them.

This is a possibility Smith seriously considers. He imagines that the poor man's son is rather inclined to treat his discontent as temporary, which "in time of sickness or low spirits is familiar to every man". But when "in better health and in better humour, … [he] … will never fail to regard" … ["those great objects of human desire"] … "under a more agreeable aspect" (2001[1759]): 183). In this case, everything remains the same, and the poor man's son does not change track. Progress, here as anywhere, is not inevitable. Advancement requires an extra effort. The advancement of knowledge requires that the poor man's son is ready to abandon his old beliefs.

Obviously the solution to the problem the poor man's son faces lies in his person. To abandon his false beliefs, he has to abandon his tendency to excess; he has to overthrow ambition and vanity. To change his beliefs, he has to put an end to his original determination; he has to change himself.

Indeed, this is exactly what the poor man's son has done. The man who looks back in the parable is no longer the same. Whereas in his youth he is "visited with ambition," in his old age he "curses ambition". The 'new man' of the parable has abandoned his extravagant passions.

Given that his strong motives completely occupied our protagonist - Smith unequivocally states that "nature imposes upon" (2001[1759]): 183) him - the question at this point is how he manages to get rid of his delusions at all. How does he overcome the resistances of his extravagant passions?

New ideas are clearly not sufficient here. These were probably around in his youth already. To adopt new ideas, it is vital that old ones be overthrown. Thus the will is needed. In the special case of the parable, this requires changing oneself, and hence the special effort of the will to "pull off the mysterious veil of self-delusion" which covers from the poor man's son the "deformities" of his own character (2001 [1759]): 158). I contend that Smith's protagonist also experiences a crisis of identity indicative of his change. The man who applies self-scrutiny and changes his character must not be deficient of will, since the resistances he has to overthrow are obviously significant. Smith compares these to the challenge a surgeon faces when he does an operation on his own body (2001[1759]): 158).

It has been noted that the deliberate questioning of assumptions is one of the general principles involved in the generation of new ideas (Loasby 1996: 25). The parable also illustrates that such questioning can hardly succeed if the agent is unwilling to overthrow the resistances against change.

The poor man's son, therefore, does not change his mind by accident. Nor is his change the result of some changing conditions, I maintain, but is motivated by the will to break with the past. The parable does not suggest that the poor man's son reacts to constraints when he pulls off his delusions. He does not *have* to change, or to know. The poor man's son changes, not because this makes him better off - actually he is not as his state of crisis shows; or because he thinks as he thought in the past - actually he thinks differently now; his change is also not the result of sudden caprice, but is a manifestation of his will.

Recent commentaries of Smith's parable often discuss the challenge it constitutes for the evaluation of welfare, a theme that clearly is involved, and with which many readers might also be more familiar. In particular, the intricate relation between wealth and happiness to which Smith alludes in the parable continuously raises attention.

On this view, the parable of poor man's son is to instruct the reader - in contrast to what most people are inclined to believe nevertheless - that wealth finally does not make them happy. Smith has the poor man's son realize that he was wrong and that he has ruined his life, as Loasby (2009b: 8) vividly puts it, since "foolishly sacrificed for what, when he has got it, can afford him no real satisfaction" (Smith 2001[1759]: 182).

Hence, on this view the deception of the poor man's son is seen in his mistaken belief that wealth can make him happy. The subject errs with regard to the true properties of the object. His error is seen to find its origin in his cognitive limitations. While recent developments have made much progress by also clarifying the nature of cognitive limitations,[38] the question whether wealth can make people happy is as yet unsolved. Consequently, differing opinions on this issue persist.[39]

Besides these ongoing and important concerns for the growth of happiness, the present account of the parable has focused on the growth of knowledge. Accordingly, it has also emphasized man's desire to know.

With a view to further clarifying the case, I will emphasize additional insights it can provide.

First, I have reinterpreted the nature and the origin of the deception involved in the parable. The poor man's son as I see it is not wrong about what the objects can do for him, nor is he wrong about his desires. But he wrongly believes that these objects do not have any limits. I have argued that the origin of this belief lies in the motives of the subject. The poor man's son disregards the existence of limits. His false belief results from his insatiability and does not find its origin in the limitations of human thought as familiar interpretations routinely claim.

The advantage of taking the motives of the agent as a starting point also is that we deal with an authentic agent whose motives are 'true', in the sense that they belong to him. This is contrary to familiar interpretations in which cognition comes first and in which false motives as a consequence of false beliefs are assumed. The poor man's son's "desire for wealth... [is] deluded", as it was recently put from the perspective of the history of economic thought (Brewer 2009). Modern accounts in this tradition routinely refer to agents who do not know, or fail "to maximize their 'true' utility" (Fleurbaey 2009: 1059).

[38] Behavioral accounts have further specified the nature of this false belief, for example in the form of a forecasting error regarding the transitoriness of happiness that can be derived from riches (Ashraf et al. 2005) or in the form of a focused illusion (Kahneman et al. 2006).

[39] To wit, the recent critique of the Easterlin paradox by Betsey Stevenson and Justin Wolfers.

As has been repeatedly pointed out, this conjecture is problematic. After all, who is in the position to know better than the agent what his true wishes are? My argument does not run into these difficulties. The ambitious agent of my interpretation is truly voracious. His avarice is not seen as the result of mistaken beliefs, but as a very real trait of his personality.

I am coming to the lesson to be taken from the present re-interpretation. That the agent is truly ambitious does not mean that he has to remain ambitious throughout his life. That his motives are authentic does not mean that they are also consistent. In fact, a man does not have to stay ambitious and vain throughout his life. He does not have to remain the person in his old days that he was in his youth. It is indicative, I think, that Smith has the poor man's son of his parable "curse ambition" but not wealth. In addition to the assumption of stable agents essential to modern welfare economics, the present account also brings out the dynamic character of human motivation.

The growth of knowledge is essential to economic development. I therefore think that it makes sense to analyze the hindrances in the way of knowledge growth. As has been shown, motives can hinder this growth. A better understanding of human motives is therefore essential.

My above analysis focused on knowledge change. In addition, it has been suggested that a person cannot only challenge and change his knowledge, but also his motives. I submit that the argument presented in relation with knowledge change can be extended to motivational change.

Motives are themselves patterns that give comfort. His motives have guided the poor man's son well. One would therefore not expect him to be a candidate for change. However, giving comfort is not equivalent to change. Change requires the will to overturn resistance. In case of the parable, it is the resistance of the 'old man' who does not want to change. To create a 'new man' - to establish a new identity - the poor man's son must also be willing to get rid of the 'old man'; he must be willing to abandon his old motives and cope with the discomfort (of crisis and depression) caused by change.

It is true that neither Schumpeter nor Smith sought to develop the idea of changing motives; nonetheless, this is not incompatible with their theories. It might reasonably be asked whether the idea is realistic at all. Can a man challenge his own success? Can he change himself?

I think that recent developments can be brought in to strengthen the case. A growing breed of poor men's sons (as 'stylized type') actually is occurring.

Personalities such as Bill Gates, Warren Buffet and Nicolas Berggruen are cases in point. They are considered among the richest men in the world thanks to the big fortunes they have made by carrying out economic innovations. They have also become well known for having given away the greatest part of their fortunes to improve the administration and growth of public wealth, and for so leading the means of production into new, public instead of private, channels of production.

Their different visions as to how and what there actually is to do notwithstanding, Berggruen has more political visions for reform of the political system than Gates, who is focused on technological advancements in care and health, what makes them interesting for us here is that these visions for a better society are carried out by men who perfectly corresponded to the picture of the Schumpeterian entrepreneur. This was before they changed and started to look out for new combinations stimulated by some discontent or crisis. Before they were able to establish new ideas, they had to abandon some of the old. Substituting public for private wealth, I believe, also requires that at least some personal change is involved.

As the fortunes that are transferred into these new channels are ever more important and ever growing, thanks to the swarm like imitation by other entrepreneurs with big fortunes around the globe and in many instances have become more important than the public money that is distributed by national and international institutions, these new men - and the new ideas and institutions and technologies they have chosen - fulfil many functions that have previously been fulfilled by the state, with ensuing consequences for the future development of the social and economic system. The relevance of the parable therefore to understand recent developments seems obvious to me.

The reader familiar with the parable of the poor man's son may also find that the parable I have presented is not Smith's but my own.

A parable is a didactic story told for instructing the reader about some truths in life. A parable normally contains different layers the truths of which are to be revealed by the reader. In contrast to the presentation of plain facts, a parable is never straightforward. It is also always an invitation to the reader to explore further perspectives. I have used the parable to accentuate Smith's deep understanding of motives in relation with the growth of knowledge. I do not make any claims that this corresponds to Smith's intentions when using the story.

Smith's parable is embedded in an important chapter of *The Theory of Moral Sentiments* in which he does not dwell on personal change, but instead emphasizes the beneficial social consequences of ambition, vanity and greed. The following passage is generally seen as indicative of Smith's view of human progress depending on the harnessing of strong motives. These strong motives and the "deception" ensuing "first prompted… [man] … to cultivate the ground, to build houses, to found cities and commonwealths, and to invent and improve all the sciences and arts, which ennoble and embellish human life; which have entirely changed the whole face of the globe, have turned the rude forests of nature into agreeable and fertile plains, and made the trackless and barren ocean a new fund of subsistence, and the great high road of communication to the different nations of the earth" (Smith 2001[1759]: 183-4). Harnessing these strong motives is promising, according to Smith, since, in combination with social influences reinforced by a system of justice, we can expect them to produce a tolerably satisfactory situation and an environment which encourages economic development.

Here is not the time and space to discuss this particular variant of what has later been dubbed 'the argument of the invisible hand.' It should only be mentioned that the idea that strong motives can be used by society to produce economic development was important also to Schumpeter's theory of innovation.

3 Comparison

My discussion of Smith's theory of knowledge confirms Schumpeter's claim on the significance of the will. The poor man's son as discussed here shows that change importantly requires the removal of resistances. On this view, Schumpeter's claim, therefore, does not only apply to economic development, but also to the foundational level of human nature.

Furthermore, the example of the poor man's son suggests that the advancement of understanding can sometimes require not only will, but extreme will. Indeed, if the advancement of understanding requires changing oneself, extensive effort is needed. It is well known that the hardest obstacles to overcome are often within oneself. This is why people generally tend to resist changing their own views (Patalano 2007; Boulding 1956). Resistance will even get stronger when personal change is required.

My discussion of the poor man's son also enriches Schumpeter's claim in two ways:

First, it provides a rationale for personal change and, by doing so, extends our insights into the motivational structure of the agent, since to understand better, and this is implied in the story of the poor man's son as emphasized here, it can be necessary to change oneself. To improve his understanding, the poor man's son has to throw off ambition, vanity and greed, i.e. the "delusions of self-love".

Second, it reveals that the will can manifest itself in various qualities and degrees of sophistication. Whereas Schumpeter's entrepreneur is fully occupied by his instincts, the poor man's son is also moved by reason and reflection. Whereas the entrepreneur is excessive, as Schumpeter reminds us when he depicts him as the "most egoistical of all", the poor man's son is able to correct his deepest inclinations. This also implies that '"unbound Prometheus' of modern capitalism" is more complex than the "almost self-evident 'stylized fact'" of the "profit motivated innovator" typically assumes (Dosi et al. 2006: 1110).

I think that Schumpeter's and Smith's theories should be seen as complementary. Then the story of the poor man's son could be understood as a sequel to the Schumpeterian account of the entrepreneur.[40] We would deal then with one - albeit evolving - agent and not with two, since a man is not the same at the beginning and at the end of his endeavor.

At the beginning, both agents are much alike: they both are determined to realize their vision. They both aspire for success, power and wealth, and they are ignorant of any bounds, since they are excessively concerned with their own person. At the end, however, these two agents could not be more different.

---

[40] In contrast to Schumpeter, who assumed that the weakening of strong motives by rationalist criticism (1966[1919]: 34) would lead to the end of entrepreneurial action, I do not believe that this is necessarily the case. If personal change happens at all, it would rather render obsolete a particular type of entrepreneur, namely the entrepreneur Schumpeter portrayed.

In contrast to the protagonist of the parable, Schumpeter does not consider that the determination of his agent ever comes to an end. If it does, this heralds his soon death in his view. But, after all, as Smith's parable also suggests, the entrepreneur is not forbidden to be disturbed in the mind and to realize the obsolescence of his beliefs with regard to approval, position and wealth. He also is not forbidden to get rid of what he considers obsolete, and hence to change his mind and motives.

So, considering that there is a sequel of events, and that agents can change and that at the end the agent is no longer the same as he was at the beginning, the parable leads us to think about the possibility that the agent finally has got enough. The poor man's son suggests that limits in the end do exist after all - albeit these are not necessarily identical with the existing limits of society and morality.

Under these circumstances, we may also ask whether other motives - motives less strong and eventually 'higher' than ambition, vanity and greed — could do the job that Schumpeter writes about at all, namely the job that has been particularly emphasized here: breaking with the resistance of the past.

I think that this finally has to remain an open question. Schumpeter, for, example is not sure either whether greed, i.e. excessive monetary compensation is needed as an assistant for the will to break with the past. He thinks that other motives (and distinctions) can also do in principle, but he also thinks that we need to think about it more specifically. In particular, he maintains that curiosity and social ambition do not have to involve the acquisition of wealth. "Pecuniary gain is indeed a very accurate expression of success, especially of relative success, and from the standpoint of the man who strives for it, it has the additional advantage obeying an objective fact and largely independent of the opinion of others. . . .Nevertheless it is true that entrepreneurial motives may in principle be taken care of by other social arrangements not involving private gain from economic innovation" (Schumpeter 1963: 94).

More recently, Dosi et al. (2006), in his empirical study, also objects that excessive profit orientation brings about more - not to mention better - innovation. It should also be remembered that the poor man's son as discussed here is not about abandoning the profit motive altogether – it is about abandoning its excesses.

4 Conclusion

There have been continuous endeavors to understand better the conditions and principles of the new, but it is probably fair to say that as yet the state of cognitive sciences has not significantly advanced (Denzau and North 1994: 21) on this theme and that "the process of origination of new combinations . . . necessarily [remains] something of a mystery" (Loasby 1996: 26).

Against this background, the present paper has suggested that we can further advance our understanding if we take a closer look at the motivational foundations of the subject: the agents of change. I have shown that our insights into cognitive processes involved in the development of ideas are not sufficient and should be completed by a thorough analysis of the human will. I have argued that change involves the special effort of the will to overcome resistance. Hence, ideas as driving forces of economic and social development (as Potts puts it, for example, 2008) alone will not do. As Schumpeter understands, for change and innovation "New Men" (1963[1934]) are needed.

The discussion of historical accounts also adds greater definition to the processes of change and innovation: destruction is more real and I understand less mystical than creation, and so is the will. The will cannot be taken for granted, but it develops, if at all, by countering the opposition to the resistances in the way of change.

As a result, processes of change of ideologies, institutions, and technologies also involve a volitional element and this implies it is also not solely a result of error or history. The agents of change do not simply adapt to their environment when they break with the past. This resonates with the standard appreciation of change as path dependent, in contradistinction to path determined (cf. for example Hermann-Pillath 2002). I have particularly brought the active element to the fore which I consider to be the most important point in relation with change that Schumpeter actually made.

As has been shown in my discussion of the mental foundations, the existence of crisis or discontent, in Smith's wording "pain" or "difficulty," is not sufficient to bring about change, their obvious significance for the stimulation of human curiosity and the search for the new notwithstanding. Else crisis would automatically be followed by renewal. As can be readily observed, this is not always the case. I think that the open ended character of the story of the poor man's son also points in this direction. We do not know what he will do next. Hence we do not know the new combinations he will perhaps find worth carrying out. From the perspective that has been emphasized here, the sequel of the story will depend not only on the capacity of the poor man's son, naturally scarce and unequally distributed among men and domains, to discover or carry out some new combinations, but also on his will to get rid of old ones and, if need be, to also change himself as a person.

The possibility of agents to change their motives - and hence virtually also change themselves - is important because it can provide a further prospective for change and development, a prospective that should not be ruled out given that a reverse causation exists according to which agents are not only affected by their ideas, institutions and technologies but also affect and choose their ideas, institutions and technologies. My analysis of the poor man's son has recalled this neglected possibility of change and development depending on the person.

The present focus on the person and personal change should be seen as proffering additional options for the further development of social arrangements. It acknowledges that motives are effective movers, but can also distort and hinder change and economic development. As both Smith and Schumpeter suggested, strong motives should certainly be used if possible, but it seems reasonable to consider as well correcting them in case of obsolescence or if they no longer fulfil their function.

The accounts of motivation discussed here go far beyond standard views in economics which have not been conceived to deal with the complexity and change of individual motives. Evolutionary economics with its focus on new developments and change holds an adequate sensor to capture their importance.

References

Arthur WB (1994) Increasing returns and path dependence in the economy. Michigan University Press, Ann Arbor

Ashraf N, Camerer CF, Loewenstein G (2005) Adam Smith, behavioral economist. J Econ Perspect 19(3):131–145

Augier M, Kreiner K (2000) Rationality, imagination and intelligence: some boundaries in human decision-making. Ind Corp Change 9(4):659–681

Blume LE, Easley D (2008) Rationality. The New Palgrave Dictionary of Economics. Second Edition. In: Durlauf SN, Blume LE (eds) Palgrave Macmillan [the new palgrave dictionary of economics online. Palgrave Macmillan. Accessed on 17 April 2009 at http://www.dictionaryofeconomics.com/article?id=pde2008_R000277 doi:10.1057/9780230226203.1390 (available via http://dx.doi.org/)

Boulding K (1956) The image. Knowledge in life and society. The University of Michigan Press, Ann Arbor, Paperbacks

Bowles S (1998) Endogenous preferences: the cultural consequences of markets and other economic institutions. J Econ Lit 36(1):75–111

Brewer A (2009) On the other (invisible) hand. Hist Polit Econ 41(3):519–544

Campbell RH, Skinner AS (1997[1976]) General introduction into Adam Smith's wealth of nations [1776]. In: Campbell RH, Skinner AS (eds) vol 1. Clarendon Press Oxford

Denzau AT, North DC (1994) Shared mental models. Ideologies and institutions. Kyklos 47 (1):3–31

Denzau AT, North DC, Roy RK (2007) Shared mental models: a postscript. In: Roy RK, Denazu AT, Willett TD (eds) Neoliberalism: national and regional experiments with global ideas, pp 14–25

Dopfer K (2004) The economic agent as rule maker and rule user: Homo Sapiens Oeconomicus. J Evol Econ 14:177–195

Dopfer K (2006) The origins of meso economics: Schumpeter's legacy, papers on economics and evolution, Max Planck Institute of Economics, # 0610

Dopfer K, Potts J (2008) The general theory of economic evolution. Routledge, London

Dosi G, Marengo L, Pasquali C (2006) How much should society fuel the greed of innovators? on the relations between appropriability, opportunities and rates of innovation. Res Policy 35:1110–1121

Elster J (1989) Solomonic judgments: studies in the limitations of rationality. MSH and Cambridge University Press, Paris et al.

Festré A, Lazaric N (2004) Routine, creativity and leadership in Schumpeter and von Mises' analyses of economic change: a new look at recent debates on routines. Paper presented at AISPE conference on "economics and institutions: contribution from the history of economics" (30 Sept–2 Oct). Palermo-Italy

Fleurbaey, M (2009) Beyond the GDP: The quest for a measure of social welfare. J Econ Lit XLVII(4):1029–75

Gerschlager C (1996) Konturen der Entgrenzung. Die Ökonomie des Neuen in Thomas Hobbes, Francis Bacon und Joseph A. Schumpeter. Metropolis, Marburg

Giocoli N (2003) Modeling rational agents. From interwar economics to early modern game theory. Cheltenham and Northampton, Edward Elgar

Grebel T (2008) Technological change. A microeconomic approach to the creation of knowledge. May. Presented at EAEPE University of Rome [published in Structural Change and Economic Dynamics May 2009]
Hermann-Pillath C (2002) Grundriss der Evolutionsökonomik. Fink Verlag, Munich. http://wga.dmz.uni-wh.de/wiwi/html/default/chph-5jvj37.en.html. Accessed 8 April 2010
Kahneman D, Krueger AB, Schkade D, Schwarz N, Stone AA (2006) Would you be happier if you were richer? A focusing illusion. Science 312:1908–1920
Keynes JM (2010[1936]) The general theory of employment, interest and money. E-books@ Adelaide. The University of Adelaide Libary
Klamer A (1989) A conversation with Amartya Sen. J Econ Perspect 3(1):135–50
Kurz H (2006) Whither the history of economic thought? Going nowhere rather slowly? Eur J Hist Econ Thought 13(4):463–88
Loasby B (2003) Connecting principles, new combinations, and routines: reflections inspired by Schumpeter and Smith. In: Metcalfe JS, Cantner U (eds) Change, transformation and development. Physica Verlag, Heidelberg and New York, pp 169–182
Loasby BJ (1994) George Lennox Sharman Shackle 1903-1992. Proc Br Acad 84:505–527
Loasby B (1996) The imagined, deemed possible. In: Helmstädter E, Perlman M (eds) Behavioural norms, technological progress, and economic dynamics: studies in Schumpeterian economics. University of Michigan Press, Ann Arbor, pp 17–32
Loasby BJ (2003) Organization and the human mind. In: Cognitivism and innovation in economics. Two Lectures delivered at the University of Milano-Bicocca and at the Bocconi University on 13–14 October 2003, pp 25–50. Accessed at http://dipeco.economia.unimib.it/pdf/pubblicazioni
Loasby BJ (2005) Entrepreneurship, evolution and the human mind. In: Evolutionary Economics Group (ed) Papers on economics and evolution. MPI, Jena
Loasby BJ (2009a) Motivation, innovation and co-ordination. In: Cantner U et al (eds) Schumpeterian perspectives on innovation, competition and growth. Springer, Berlin Heidelberg, pp 37–54
Loasby BJ (2009b) Imagination, illusion and delusion. Sceme Working Paper 27
March JG (1995) The future, disposable organization and the rigidities of imagination. Organization 2:427–440
McCraw TK (2006) Schumpeter's business cycles as business history. Bus Hist Rev 80:231–261
Morrison K, Potts J (2008) Toward behavioural innovation economics - Heuristics and biases in choice under novelty. Discussion Paper. No. 379, The University of Queensland, pp 1–21, http://www.uq.edu.au/economics/abstract/379.pdf. Accessed 30 Nov. 2009
Nelson RR, Sampat BN (2001) Making sense of institutions as a factor shaping economic performance. J Econ Behav Organ 44:31–54
Nelson RR, Winter SG (1982) An evolutionary theory of economic change. The Belknap Press of Harvard University Press, Cambridge, Massachusetts
North D (1999[1990]) Institutions, institutional change and economic performance. Cambridge University Press, Cambridge
Patalano R (2007) Images and economic behaviour. Discussion Paper 2, University of Naples, pp 1–12
Pavitt K (1998) Technologies, products and organization in the innovating firm: what Adam Smith tells us and Joseph Schumpeter doesn't. Ind Corp Change 7(13):433–52
Pesciarelli E (1989) Smith, Bentham, and the development of contrasting ideas on entrepreneurship. Hist Polit Econ 21.3: 521–36
Potts J (2008) Economic evolution, identity dynamics and cultural science. Cultural Science 1(2). http://cultural-science.org/journal/index.php/culturalscience/article/viewArticle/16/54
Raphel DD, Skinner AS (1980[1795]) General introduction into Adam Smith's essays on philosophical subjects. In: Wightman WPD, Bryce JC (eds) Essays on philosophical subjects. Oxford University Press, Oxford, pp 1–27

Rothschild E, Sen A (2006) Adam Smith's economics. In: Knud Haakonssen (ed) The Cambridge Companion to Adam Smith. Cambridge University Press, Cambridge et al., pp 319–65

Schumpeter JA (1911–12) Theorie der wirtschaftlichen Entwicklung, 1st edn. Duncker & Humblot, Leipzig

Schumpeter JA (1926) Theorie der wirtschaftlichen Entwicklung, 2nd edn. Duncker & Humblot, Munich and Leipzig

Schumpeter JA (1942[1975]) Capitalism, socialism and democracy harper perennial, New York

Schumpeter JA (1954[1986]) History of economic analysis. Routledge, London

Schumpeter JA (1963[1934]) The theory of economic development. An inquiry into profits, capital, credit, interest, and the business cycle, 3rd edn. Translated by Redvers Opie. Oxford University Press, New York

Schumpeter JA (1966[1919]) The sociology of imperialisms. In: Hoselitz B (ed) Imperialisms and social classes. Two Essays by Joseph Schumpeter, 9th edn. Transl. by Heinz Norden. Meridian Books, Cleveland and New York, pp 3–98

Shackle GLS (1967[2010]) The years of high theory. Invention and tradition in economic thought 1926–39. Cambridge University Press, Cambridge et al.

Shackle GLS (2009[1970]) Expectation, enterprise and profit. The theory of the firm. Allen & Unwin, London

Shackle GLS (1972) Epistemics and economics. Cambridge University Press, Cambridge

Skinner AS (1996) A system of social science: papers relating to Adam Smith, 2nd edn. Clarendon Press, Oxford

Smith A (1980[1795]) The principles which lead and direct philosophical enquiries, illustrated by the history of astronomy. In: Wightman WPD, Bryce, JC (eds) Essays on philosophical subjects, Oxford University Press, Oxford, pp 32–105

Smith A (1983[1762]) Lectures on rhetoric and belles lettres, Glasgow Edition, vol.4, Bryce, JC (ed) Clarendon Press Oxford. THE ONLINE LIBRARY OF LIBERTY Liberty Fund, Inc. 2005. At http://oll.libertyfund.org/Home3/index.php

Smith A (1997[1776]) The wealth of nations. In: Campbell RH, Skinner AS (eds) Glasgow Edition, vol 1. Oxford University Press, Oxford

Smith A (2001[1759]) The theory of moral sentiments. Glasgow edition. Raphael DD, Macfie AL (eds) Oxford University Press, Oxford

Vanberg VJ (2002) Rational choice vs. program-based behaviour. Ration Soc 14(1):7–54

Vromen J (2001) The human agent in evolutionary economics. In: Laurent J, Nightingale J (eds) Darwinism and evolutionary economics. Edward Elgar, Cheltenham, pp 184–208

Network Evolution in Basic Science

Thomas Grebel

Abstract Under true uncertainty, as is the case in basic science, researchers need apply generic rules to overcome the imponderability of inventive activities. How such rule-behavior may influence the evolution of networks is modeled in this paper. An empirical survey on medical researchers describes their research context and serves as a motivation to construct a percolation model that illustrates the expected structural evolution of networks. Leaders and followers in a new research field induce a stylized structural evolution of networks.

1 Introduction

Knowledge produced in basic research usually is far from applicable. Even in applied research, it takes decades before a marketable technology is accomplished. Consequently, economic returns to potential technological actualizations are negligible drivers of research efforts in basic science. The chances to appropriate possible rents are low (Arrow 1962). The incentive that guides research behavior in medicine is twofold: first, research is an integral and obligatory part of the medical education system; second, it is the foundation of a researcher's scientific reputation which is indicated by his stock of publications, whereas each publication stands for being awarded the *priority in discovery*. Hence, each publication is the reward for scientific research efforts (Dasgupta and David 1994).

How does one achieve this reward? Because of true uncertainty, there is no operant rule that would determine optimal behavior. Researchers need to apply generic rules, rules that may lead to different actualizations, and to undesired ones (Dopfer 2004). To overcome the lack of sufficient information to act optimally,

Reprinted from Journal of Evolutionary Economics 22(3), 443–457, Springer (2012).

T. Grebel (✉)
Economics Department, Friedrich-Schiller-University, Jena, Germany
e-mail: thomas.grebel@uni-jena.de

G. Buenstorf et al. (eds.), *The Two Sides of Innovation*, Economic Complexity and Evolution, DOI 10.1007/978-3-319-01496-8_3,

there are several conceivable generic rules researchers apply: researchers may just look out for new research fields, as the objective is to generate something novel; or, they could cooperate with others, since cooperation reportedly increases the odds of successful research outcomes. Given cooperation is beneficial in any case, there are also different underlying reasons why researchers may cooperate. Institutional frames, for example, put researchers in different contexts which, in turn, influence individual behavior.

A further aspect, aside from contextual aspects, questions the individual decision-making process, as to why individuals decide to join a certain network. In a largely exploited research field, latecomers would rather refrain from starting to work in the new field because the chances to generate something novel are low. However, if the new field turns out to become a new research paradigm substituting for the current one, researchers may join a network opposite to their initial beliefs. This is a phenomenon discussed by Wärneryd (2008) and Tarde (1890), who distinguish between *logical* and *extralogical* behavior. Logical, in this sense, means that researchers deliberately join a network because they are convinced of both the new research field and the usefulness of the cooperation partners. They can be considered as leaders. By contrast, extralogical would refer to researchers who just imitate the behavior of others, join a network and follow along so as to work in the same research field. They behave as if they act against their own beliefs, but do not see other options. They follow the leaders.[1] Though a researcher may not be convinced of the usefulness of a research topic, he has the desire to take a share in the reward, that is, publishing with other researchers in the network. Thus, mimicking the behavior of others seems beneficial.[2]

The objective of this paper is to construct a simple model of generic network behavior in basic science. It builds on the actualization of generic rules, rules that are applied under strong uncertainty. The methodological frame employed draws on the work by Dopfer and Potts (2008) and Dopfer (2004). To distinguish between a *generic* and an *operant level* in economic analysis. Whereas the latter focuses on operations on commodities, as in traditional economics, the former investigates rules and their actualizations. A percolation model will put those building blocks together. To provide an empirical motivation, a survey on medical researchers in Germany is illustrated. Step by step the generic-rule-based behavior will be incorporated into the percolation model. The percolation model itself is supplemented by a Monte-Carlo-Simulation to check the robustness of the results. In Section 2, the survey results will give a rough picture of the context of researchers. Though the survey results will not be sufficient to back all generic rules assumed in the paper, some contextual elements will be supported, which resonate with some characteristics of percolation theory. An understanding of generic elements in network

[1] According to Wärneryd (2008), the view by Tarde (1890) has long been neglected. Tarde delivers a theory on economic psychology, which matches quite well the dichotomy of a Schumpeterian entrepreneur versus a non-entrepreneur in research in basic science: a proactive versus a less active researcher.

[2] See also Banerjee (1992), Ellison and Fudenberg (1995) and Schlag (1998).

evolution will be provided in Section 3. The actual model of network evolution is presented in Section 4. Section 5 summarizes the structural phases of the model and concludes.

2 The Organization of Medical Research in Germany

The results of a survey on the organizational structure of research at German university clinics and medical research institutes, performed at the Friedrich-Schiller University in Jena,[3] illustrate the research context of medical researchers in Germany.

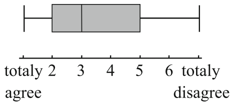
Survey Description. In summer 2008, an online questionnaire was sent out to around 18,000 researchers at medical clinics and research institutes in Germany. The response rate was about 10% (1,747 returned questionnaires). The sample comprised 292 researchers from theoretical institutes, 1,140 from clinical institutes with beds and 312 from clinical institutes without beds. All respondents indicated to do research in medical basic science. They came from all hierarchical levels of their organization, with 33% of researchers having achieved a habilitation,[4] of which 17% were clinical directors, 49% assistant medical directors with a Ph.D., and 19% assistant physicians having accomplished their medical studies. The topics covered in this survey were on the respondents' educational profile, their work load, their autonomy in research activities, the influence of research group leaders and clinic directors on their research, multidisciplinarity, composition of research groups. Detailed results can be found in Cantner et al. (2009).

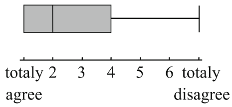
Organizational Context and Generic Behavioral Rules. The organizational structure of a university clinic and medical research institutes is hierarchical and thus aligned to the needs of patient care. There is a director responsible for the entire clinic, followed by directors of the various medical divisions such as internal medicine, pediatrics and geriatrics. A further hierarchical step below, assistant medical directors lead medical groups, which mainly consist of assistant physicians. One may assume that the hierarchical structure necessary to ensure efficient patient care also reflects the organizational structure of research.

According to the survey results, however, the research context seems to be quite flexible, despite its hierarchical organizational structure. The respondents were asked about their research autonomy, whether they are in the position to make independent individual decisions, in the first place, or whether they have to join certain networks as it coincides with the position in the respective organization. The distribution of answers given by respondents is shown below:

[3] The research project was named "Innoheart" and financed by the "Bundesministerium für Bildung und Forschung", the German ministry of research and education.

[4] The Habilitation is a degree in the German academic system usually made to become professors.

1. In my research work, I work independently from others. 

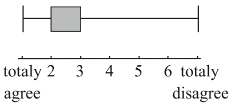
2. I decide on the subject matter of my research all by myself. 

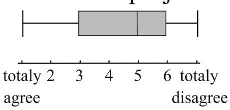
The Box–Whisker–Plot next to the questions provide a rough idea about the distribution of the respondents' answers measured by a 7-point Likert-scale from 1 (totaly agree) to 7 (totaly disagree). Answers to question 1 suggest that researchers have enough room to make their own decisions and can work rather independently from others. The median of the answers is 3. Question 2 inquires about the influence a researcher has on the contents of his research work. The most common answer was 2, which signifies that researchers tend to assume themselves to be fairly autonomous in their research. Though these answers do not explain the actual behavior of researchers, they suggest that there is room for individual decision making by researchers. With regard to our modeling intentions, the following assumption is made:

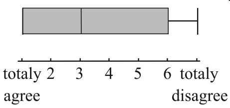
Assumption (1) Medical researchers can choose their research topic autonomously

This assumption implies that researchers need to apply behavioral rules, in order to determine their behavioral actions. They know that the outcome is uncertain and therefore have to undergo a trial-and-error process.

Concerning the emergence of networks, question 3, 4 and 5 were asked to find out where the members of a research group stem from:

3. Potential researchers get informed about new research projects by personal contacts. 

4. Research project participants are recruited via advertised bidding. 

5. I work in a research group which is embedded exclusively in our organization. 

On average, most of the potential research group members are informed about new research projects by personal contact. Participants recruited by job advertisements are rare. Answers to question 5 show that, on average, research groups consist of local researchers belonging to the same organization, although it has to be conceded that there is a wide dispersion of answers. Nevertheless, by additionally taking into account Polanyi (1958) and von Hippel (1994), it seems legitimate to interpret these results as a support for the importance of local proximity for cooperative behavior. Hence, a further assumption is:

Assumption (2) Local proximity enhances cooperation.

3 Generic Elements in Network Evolution

The creation of knowledge is a generic process with a multitude of possible unknown outcomes. Dopfer and Potts (2008) label the economic analysis of such kind of processes a generic analysis. Researchers apply *rules* to motivate their behavior. Although there are some *operant* rules that lead to unique and ex ante known consequences (actualizations), such as any optimization rule would do, there are a lot of so-called *generic rules* that can lead to a multitude of possible actualizations which are unknown ex ante.

With respect to basic research the scope of generic rules that have to be applied is quite large. As already mentioned, there is no unique operant rule that tells us how to generate positive research results. Instead, researchers apply generic rules and look out for new research fields, take an effort in doing research in it, and cooperate with other researchers.

In the following, the elements addressed so far will be compiled in order to build a percolation model. Percolation theory is used because it offers a lot of characteristics which allow combining the idea of generic rules applied in basic research behavior to the context of researchers in medical research.

Percolation Theory. There are a multitude of techniques and methods mentioned in the literature to model network behavior and network evolution. Erdős and Rényi (1959, 1960, 1961) give a general overview, discussing random graph theory, study general properties such as the connectedness of a graph, and the circumstances under which triads or subgraphs emerge. Bollobàs (1985) specifically asks questions about the probabilities and the distribution of subgraphs; and investigates what happens if the number of actors goes to infinity. For the purpose of this paper, however, the simplest approach appears to be percolation theory.

The inherent characteristics of percolation theory,[5] such as the role of local proximity or the so-called percolation threshold, offer a simple way to describe the

[5] In physics, percolation theory describes the penetration of a medium by a fluid. Two sites are connected if both are permeable to the fluid, such as the pores of a coffee filter. Only if a sufficient number of open pores are connected will the fluid permeate the whole medium.

cooperative behavior of researchers in basic science. As local proximity matters, the percolation threshold, usually labeled p_c,[6] allows us to model a symmetry break in research behavior with respect to the role of knowledge diffusion. The percolation threshold denotes the moment at which a lattice-spanning network (cluster) emerges (Albert and Barabàsi 2002). Below this threshold, $p < p_c$, there primarily are isolated networks.[7] If there are only isolated networks on the lattice, it is less likely that members doing research within the boundaries of such disjunct network run the risk of losing the scientific appropriability of their new insights, since the diffusion of knowledge beyond the limits of the networks is restricted. However, if a lattice-spanning network emerges, the incentive to generate knowledge and cooperate with others should decline as the probability to keep knowledge local declines.

The basic idea of a percolation approach is the following: consider a square lattice of dimension d, each square representing a researcher embedded in a neighborhood. The neighborhood of an actor can be described by the von-Neumann neighborhood. It depicts the four neighbors of an actor in the north, south, west and east of the square that identifies an actor.[8] The embedding into a lattice allows us to define distances and model locality. A site on the lattice can be considered as connected to a neighboring site, if both sites are *occupied* or *open*. A site can as well be interpreted as a researcher, and an open site as a researcher having received information about a new research field. In case two neighboring sites are open, then they are connected, or analogously, two neighboring researchers (von-Neumann neighborhood) start to cooperate after having received the information of an upcoming new research field.

4 Generic Network Evolution

In this section, a model of network evolution is presented, which exclusively builds on generic behavioral rules. Step by step, the respective generic rules will be implemented. As a starting point, a lattice of finite size is used, with each site

[6] There are many examples where percolation models are used in economics: product diffusion models are the most prominent ones (Hohnisch et al. 2008). In marketing, it is used for evaluating a product's possible adoption process (Chandrasekaran and Tellis 2007) and thus incorporates theories about the demand side such as consumer preferences (Witt 2001); others take path dependence (David 1985) and communication patterns (Ratna et al. 2008) into account. Social learning processes are addressed by Young (2007), underlying externalities by Allen (1982), and the emergence of different market structures is investigated using percolation theory (Solomon et al. 2000).

[7] Stauffer and Aharony (1994) provide a detailed overview on possible applications of percolation theory. In this respect, random graph theory and percolation theory can be used interchangeably (Albert and Barabàsi 2002).

[8] A Moore-neighborhood would add the directions: north-west, south-west, north-east and south-east. However, it will not change the basic results of the model.

representing an actor of a given community in medicine. In Section 4.1, a naive behavior is assumed with researchers applying just two simple generic rules: If there is a new emerging research field, they start to work in it and cooperate with nearby informed researchers. In Section 4.2, further generic rules are added discriminating *leading* from *following* researchers: researchers who deliberately decide in favor of the new research field and cooperation options, and researchers who just follow others by *imitating* their behavior.

4.1 Basic Model

Let the size of the lattice be of dimension d so that the number of researchers n amounts to d^2. The neighborhood of a researcher is defined as a von-Neumann-neighborhood.[9] Suppose a new research topic is coming up. While the information about this new field is diffusing, researchers who receive this information start to do research in it. The diffusion of the information simply follows a stereotypical diffusion curve.[10] Hence, the propensity of a researcher to start working in the new field is assumed to be p, which represents generic rule (1):

Generic rule (1) Having received information about the new research topic, researchers start working in it.

If $p \rightarrow 1$, all n actors eventually do research in the new field. According to assumption (2), local proximity enhances cooperation. This assumption manifests in the following generic rule:

Generic rule (2) Cooperate with an equivalently interested and informed researcher in your neighborhood.

To model cooperation, we make use of the innate quality of percolation theory. If two researchers belong to the same von-Neumann neighborhood and both are informed by the new research topic, they start cooperating and form a network tie.

---

[9] There are also other options to define neighborhoods using a Cayley tree or a von Moore-neighborhood. Such concepts can be implemented easily without changing the general propositions of this paper.

[10] Grebel (2004) gives an example how to implement such a bimodal diffusion process: the diffusion of knowledge over time and the diffusion of entrepreneurial actions.

Hence, the probability that a researcher stays a non-cooperating single researcher is: $(1 - p)^4$.

First simulation results are illustrated in Fig. 1. Each simulation run is an artefact because the diffusion process is a random process. Therefore, a Monte-Carlo simulation is applied in order to show the average expected emerging network structures subject to *p*.[11]

The Monte-Carlo simulation comprises $N = 2{,}500$ as a fixed number of researchers and 200 simulation runs. The horizontal axis in Fig. 1 depicts the assumed fraction *p* of actors who work in the new research field. The uppermost diagram shows the number of informed researchers *IR*, which is equal to *p*. *NC* counts the number of networks dependent on *p*.[12] If everybody is informed, all researchers belong to a single network. A more detailed picture on particular quantities is given in the middle panel: dependent on *p*, the maximal size of a network (*MaxCS* (solid line)) stays low until $p = 0.3$, then it starts to increase sharply. For $p = 1$, the maximum size of the network is *N*. The minimum size of networks is $MinCS = 1$ (line of long dashes) up to $p = 0.8$, which is due to the fact that it is still possible to observe isolated researchers. If the fraction *p* of researchers who get informed and switch to the new field increases even further, the minimum size steeply increases with $p > 0.8$. For $p = 1$, $MinCS = MaxCS = N$. The average size of networks (*AvgCS* (line of small dashes)) follows *MinCS*, that is, the share of smaller networks in overall networks is high. The standard deviation of the network size *StdCS* remains low until $p = 0.7$, increases to its maximum at about $p = 0.85$, and eventually goes down to zero when $p = 1$. The diagram bottom illustrates the number of single researchers (*SR*) and research groups (*RG*). Research groups denote a network with at least two cooperating researchers.

All three panels depict quantities that describe structural equilibria subject to *p*, which is the number of informed researchers. In the model, *p* is an exogenous parameter, as it is intended to investigate the evolving structures of networks subject to *p*. In reality, *p* is endogenous. This parameter is crucial for the scope and shape of diffusion. If the new field does not become a new general research paradigm, *p* stays below 1 or may even decline again in the course of time. So far, the model simply suggest the expected network structures when *p* ranges from zero to one.[13]

If *p* ranges from zero to one, as suggested in Fig. 1, the curves can be divided into five phases. Phase I: the number of networks (NC, top diagram)[14] increases with *p*. NC, decomposed in the number of isolated actors (SR) and research groups (RG) in

[11] Albert and Barabàsi (2002) and Sahimi (1994) provide a formal basis for the emerging quantities produced by the model.

[12] For simplicity, two or more neighboring informed researchers are considered a network.

[13] The shape of diffusion curves usually differ and are non-linear. A time component will distort the phases in the diagrams in Fig. 1, either expanding or contracting respective phases, though they do not change the general picture.

[14] Note, NC counts single actors and research networks of size > 1.

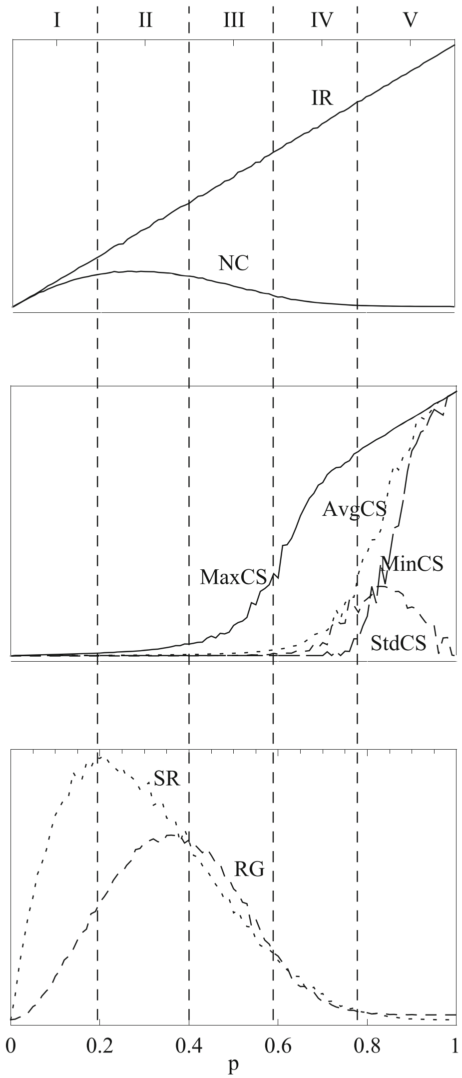


Fig. 1 Evolution of generic network structures. Monte-Carlo simulation with $N = 2{,}500$ and 200 simulation runs. *IR*: informed researchers, *NC*: number of clusters, *MaxCS*: maximal size of cluster, *AvgCS*: average size of cluster, *MinCS*: minimal size of cluster, *StdCS*: standard deviation of cluster size, *RG*: number of research groups, *SR*: single researchers

the bottom diagram, shows that the number of single researchers and research groups increase until $p = 0.2$. At the end of phase I, the difference between the number of single researchers (SR) and the number of research groups (RG) is maximal. From $p = 0.2$ on, the likelihood to observe isolated actors declines. In turn, the probability of emerging networks with size > 1 augments and exceeds the number of single actors at $p = 0.4$ (end of phase (II), bottom diagram, SR intersects RG).[15] In phase (III), the number of research groups (RG) exceeds the number of single researchers (SR). The maximum network size (*MaxCS*) remains low. At the

[15] The maximum number of networks lies at $p = 0.3$.

percolation threshold ($p = 0.593$),[16] the size of the biggest network skyrockets and a dominant network emerges (diagram in the middle). As a consequence, the number of networks must decrease because more and more researchers get linked to networks, and networks in turn get connected to other networks. In phase (V), no single researchers are observed any longer, since the minimum size of networks sharply increases. In the end of the last phase, all actors become connected. All researchers participate in the same network, given the assumption that researchers would only use the two generic (naive) rules above. As already mentioned, strategic behavior is omitted in this picture. We have not yet considered the reasons why researchers join the network. So far, we simply find the expected network formation dependent on (p).

Now, let us have a closer look at the respective phases *IV* and *V*. This kind of behavior does not appear reasonable. At the percolation threshold (beginning of phase *IV*), the probability of a lattice-spanning network is very high. This means that researchers are highly connected, which increases the risk of a loss in scientific knowledge appropriability. Knowing that every other researcher is already working in this research field, the remaining researchers should be deterred from starting to work in it as well, since a highly connective structure of a network enhances the diffusion of new knowledge. Consequently, knowledge leakages increase, and therefore new knowledge becomes available to a larger group of medical researchers in the medical community (Arrow 1962). This should affect the researcher's attitude concerning his commitment as well as the incentive to cooperate. Hitherto, no such generic rule has been implemented. To overcome this shortcoming strategic generic rules will be introduced next.

4.2 Leaders and Followers

At the percolation threshold ($p = 0.593$), a lattice-spanning network has to be expected, and knowledge created within the network can easily diffuse among community members. Therefore, the incentive to engage in this research field should go down. It should go down even further the more researchers have already been working in the new research field, since the chances of making new discoveries in an exploited field presumably decline. Therefore, it seems plausible to assume a further (strategic) generic rule researchers may apply:

Generic rule (3) The more researchers at work in a field, the less likely is one to start working in it.

[16] The percolation threshold is $p = 0.5927$ with respect to site percolation (Sahimi 1994). Compare also Newman and Ziff (2000).

In order to incorporate this rule, a simple function $f(p)$ with the following properties is used: $f(0) = 1$, $f(p_c) = 0$, $f'(p) < 0$ and $f''(p) > 0$. The function $f(p)$ is bounded between 0 and 1 with a decreasing slope at a diminishing rate. Basically, it weights the propensity of an informed researcher's decision to work in the new research field. If the researcher decides against it, he sticks to the current research paradigm, unless p keeps on increasing. In the case of the new research field becoming the dominant research paradigm, p will keep increasing even further, finally reaching $p = 1$. Then, researchers initially not favoring the new research field will have switched to the new paradigm, a behavior called *extralogical* by Wärneryd (2008) and Tarde (1890).[17]

If a new research field becomes a new research paradigm and substitutes for the previous one, certain researchers who initially had not actively decided to work in this new field (*followers*) start to imitate the behavior of others in their proximity. They start to switch to the new field and join the research network in their neighborhood. Followers switch rules contrary to their own interpretation of the information received. Initially, they may have decided against it, as they thought the field was already too crowded. But because of the increasing number of other researchers who have started doing research and cooperate in the new field, they do so as well.[18] The corresponding generic rule assumed is:

Generic rule (4)	The more actors switch to the new research field, imitate their behavior and cooperate with active researchers in the neighborhood.

Note that generic rule (4) compensates the initial decision of followers not committing to the new research field as p increases. If p approaches 1, the initial decision of reluctant researchers not to focus on the new research field is fully compensated for by generic rule (4). Certainly, this would only be the case, if the emergence of a new research paradigm is assumed to wipe out completely the old one. Researchers who are labeled followers by acting in accordance with generic rule (4) switch only if there is a network of active researchers in their proximity. If they see close colleagues working there, they start to join as well.

[17] Tarde (1890) labels this kind of behavior *extralogical* behavior. It is "(...) blind copying without consideration of utility, spurred by the feeling that the beliefs or behaviors of superiors should be copied to the best of one's ability." (Wärneryd 2008).

[18] Banerjee (1992), Ellison and Fudenberg (1995) and Schlag (1998), for example, investigate the phenomenon of mimicking rule-based behavior. In a game theoretical context, pay-offs to certain rules induce a learning process such as: pull the left arm of a two-armed bandit yields a higher pay-off. The pay-off of research behavior is not as immediate as bandit theory suggests. There are, at the most, long-term pay-offs of joining a certain research network in basic science. The mimicking behavior is rather of the type: if everybody else seems to switch to the new field, I do so, too.

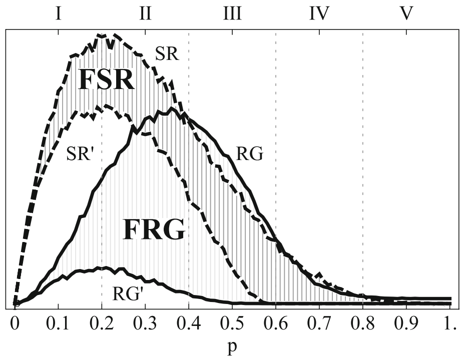


Fig. 2 Evolution of network structures: including generic rule (3). Monte-Carlo simulation wit $N = 2{,}500$ and 200 simulation runs. *RG*: number of research groups, *SR*: single researchers, *FSR*: following single researchers, *FRG*: following informed research groups

Figure 2 shows a Monte-Carlo simulation with $N = 2{,}500$ and 200 runs with the two additional generic rules (3) and (4) included. The diagram shows that with these rules included, the number of (single) researchers (*SR'*) who actually become leaders in this field is lower than the number of researchers depicted in the previous Fig. 1; and as well, fewer research groups (*RG'*) emerge compared to the case without theses rules. The dashed line of single researchers *SR* (model, without generic rule (3) and (4)) is pushed down to *SR'* (model with generic rule (3) and (4) included), and the solid curve of research groups *RG* (model without generic rule (3) and (4)) to *RG'* (with generic rule (3) included). All points on the curves indicate structural equilibria the respective model produces. It distinguishes leading researchers from following researchers: leaders deliberately decide to work in the new research field, i.e. researchers depicted by curves SR' and RG', and followers—*SR* minus *SR'* (plane *FSR*) and *RG* minus *RG'* (plane *FRG*)—who are informed but do not take the proactive decision to engage in the new research field. Hence, the two corresponding hatched planes in Fig. 2 depict the fraction of followers: single researchers and researchers in groups, respectively.

Drawing only on four simple generic rules, the model quantifies the number of networks and their topology. The methodological frame, as suggested by Dopfer and Potts (2008), Dopfer (2004) and Dopfer et al. (2004), and which is applied here, appears to be helpful in investigating emergent phenomena such as the evolution of research networks in basic science. Combining this approach with the tools supplied by percolation theory allows us to derive some general propositions to the structural evolution of researcher networks. Neither of the generic rules assumed can be considered an optimization rule. Both, leaders as well as followers, may end up in suboptimal situations, as strong uncertainty does not allow for optimal rules. Researchers know that their decisions with respect to research activities are fallible. But decisions have to be make, and generic rules become the last resort.

Table 1 Generic behavioral rules

Behavioral rules	Percolation phase				
	Phase I	Phase II	Phase III	Phase IV	Phase V
Generic					
Research incentive	High	Decreasing	Low	Nil	Nil
Leaders (logical) cooperation	Increasing	Decreasing	Low	Nil	Nil
Followers (extralogical) cooperation	Low	Low/ increasing	Increasing	Herd behavior	Complete adaptation

5 Summary and Conclusions

Instead of using an optimization approach, this paper draws on a methodological frame based on generic rules as the driving behavioral forces in the formation of research networks in basic science. A survey on the organizational context of medical researchers in basic science motivates the story. It shows that, despite the hierarchical structure in German medical clinics and institutes, researchers appear to have the chance to make their own decisions about their research activities, such as choosing their research topic and possible cooperation partners. The decisions they make are subject to strong uncertainty. This deprives researchers from operant rules that provide a deterministic chain of behavioral actions so as to achieve optimal outcomes. The attempt to optimize is futile. Instead, they have to apply generic rules. Conclusively, the emerging structure of networks cannot be considered to be the result of optimal behavior, either. Researchers have to apply generic rules in their cooperative behavior, too.

The model in this paper uses four simple generic rules serving as behavioral assumptions for the percolation model. In search of new discoveries, researchers are alert to new research fields. If a new research field emerges, researchers decide whether to focus on the new field and to cooperate with other researchers in their proximity. A crucial parameter in the decision-making process of researchers is p: the fraction of researchers at time t who are already committed to the new field. Cooperative behavior emerges from two generic sources: (a) researchers deliberately deciding in favor of the new research field and pro-actively choosing to cooperate (leaders), and (b) researchers who just follow and imitate the behavior of others (followers). Discriminating between these two kinds of generic dispositions also has an impact on the actualized evolving structure of networks. As followers do not form own networks but rather join nearby existing networks, the size of networks will increase disproportionately, whereas the number of networks declines, given p increasing. Table 1 summarizes the respective results: the incentive to start research in a new field, once a researcher has received information about it, should be positive, since a new field may suggest higher chances to discover novelties and to publish. This incentive is high in phase (I), decreases in phase (II) and fades out after phase (III) as more and more researchers exploit the field. The same holds for cooperative behavior. Leaders deliberately decide to pursue research in the new field and cooperate with nearby colleagues. Beyond the percolation threshold p_c, however, this incentive goes down, since the number of engaged researchers increases. In contrast to the *logical* behavior of leaders, followers apply a generic rule that seems to be illogical at first glance. This *extralogical* behavior is reflected by generic rule (4). With p being low, followers do not have an incentive to follow others, as they might not value the new research field. But with p increasing, more and more followers switch to the new research field and cooperate with close neighbors as they do not know any better. If p goes to one, the share of followers reaches its maximum share.

In future research work, there are several tasks to tackle. Parameter p should be endogenized. As not all emerging research fields become new paradigms, p is very unlikely to reach one. Additional generic rules should be introduced to allow for a more elaborate behavior. Optional new research fields researchers may choose from should be included. Furthermore, a link to medical markets has to be made, since medical research networks will be influenced by firm interest as soon as the exploitation of a research field becomes promising for firms that seek to innovate. Empirically, it is necessary to collect data on p and test the model results, such as the number and size of networks.

All in all, this exercise has shown that the generic rule-based approach advocated by Dopfer et al. can be useful in investigating emergent phenomena, particularly if we deal with actors who have to cope with true uncertainty such as in basic research, where the attempt to optimize is rather illusive.

References

Albert R, Barabàsi A-L (2002) Statistical mechanics of complex networks. Rev Modern Phys 74 (1):47–97

Allen B (1982) Some stochastic processes of interdependent demand and technological diffusion of an innovation exhibiting externalities among adopters. Int Econ Rev 23(3):595–608

Arrow K (1962) Economic welfare and the allocation of resources for invention. In: Nelson RR (ed) The rate and direction of inventive activity: economic and social factors. Princeton University Press, Princeton

Banerjee AV (1992) A simple model of herd behavior. Quart J Econ 107(3):797–817

Bollobàs B (1985) Random graphs. Cambridge University Press, Cambridge

Cantner U et al (2009) Chancen und Risiken minimalinvasiver, kardiochirurgischer Herzmedizin. Bundesministerium für Bildung und Forschung

Chandrasekaran D, Tellis GJ (2007) A critical review of marketing research on diffusion of new products. Review of marketing research, pp 39–80

Dasgupta P, David P (1994) Toward a new economics of science. Res Policy 23(5):487–521

David PA (1985) Clio and the economics of qwerty. Am Econ Rev 75:332–337

Dopfer K (2004) The economic agent as rule maker and rule user: homo sapiens oeconomicus. J Evol Econ 14(2):177–195

Dopfer K, Foster J, Potts J (2004) Micro-meso-macro. J Evol Econ V14(3):263–279

Dopfer K, Potts J (2008) The general theory of economic evolution. Routledge, London and New York

Ellison G, Fudenberg D (1995) Word-of-mouth communication and social learning. Quart J Econ 110(1):93–125
Erdős P, Rényi A (1959) On random graphs I. Publ Math 6:290–297
Erdős P, Rényi A (1960) The evolution of random graphs. Publ Math Inst Hung Acad Sci 5:17–61
Erdős P, Rényi A (1961) On the evolution of random graphs. Bull Inst Internat Statist 28:343–347
Grebel T (2004) Entrepreneurship—a new perspective. Routledge, London and New York
Hohnisch M, Pittnauer S, Stauffer D (2008) A percolation-based model explaining delayed takeoff in new-product diffusion. Ind Corp Change 17(5):1001–1017
Newman MEJ, Ziff RM (2000) Efficient monte carlo algorithm and high-precision results for percolation. Phys. Rev. Lett. 85(19):4104–4107
Polanyi M (1958) Personal knowledge: towards a post-critical philosophy. Routledge and Kegan Paul
Ratna N, Dray A, Perez P, Grafton Q, Newth D, Kompas T (2008) Diffusion and social networks: revisiting medical innovation with agents. In: Hassan Q-U, Michael SJ, Davidsen PI (eds) Complex decision making: theory and practice. Springer-Verlag, pp 247–268
Sahimi M (1994) Applications of percolation theory. Taylor & Francis, London
Schlag KH (1998) Why imitate, and if so, how? A boundedly rational approach to multi-armed bandits. J Econ Theory 78(1):130–156
Solomon S, Weisbuch G, de Arcangelis L, Jan N, Stauffer D (2000) Social percolation models. Phys Stat Mech Appl 277(1–2):239–247
Stauffer D, Aharony A (1994) Introduction to percolation theory, vol 2. Taylor and Francis, London
Tarde G (1890) Les lois de l'imitation. Alcan, Paris
von Hippel E (1994) Sticky Information and the locus of problem solving: implications for innovation. Manage Sci (40(4)):420–439
Witt U (2001) Special issue: learning to consume—a theory of wants and the growth of demand. J Evol Econ 11(1):23–36
Wärneryd K-E (2008) The psychological underpinnings of economics: economic psychology according to gabriel tarde. J Socio-Econ 37(5):1685–1702
Young HP (2007) Innovation diffusion in heterogeneous populations. Economics Series Working Papers 303, University of Oxford, Department of Economics

Inventor Networks in Emerging Key Technologies: Information Technology vs. Semiconductors

Holger Graf

Abstract This paper analyzes the development of the German knowledge base measured by co-classifications of patents by German inventors and relate this technological development to changes in the structure of the underlying inventor networks. The central hypothesis states that technologies that become more central to the knowledge base are also characterized by a higher connectedness of the inventor network. The theoretical considerations are exemplified in a comparative study of two patenting fields—information technology and semiconductors. It turns out that information technology shows the highest increases in patents, but only a moderate move towards the center of the knowledge base. By contrast, semiconductors develops towards a key technology, despite a moderate increase in the number of patents. The dynamic analysis of inventor networks in both fields shows an increasing connectedness and the emergence of a large component in semiconductors, but not in information technology, which is in line with the expectations.

1 Introduction

Ideas might be conceived in isolation, but research and the diffusion of its results is an interactive process in which scholars build on one another's work, collaborate in joint projects, and exchange knowledge in various ways (Andersson and Beckmann 2009). In a recent study, Wuchty et al. (2007) show that research across all fields is increasingly performed by teams. Between 1975 and 2000, average team size in patenting rose from 1.7 to 2.3 inventors per patent and the tendency towards increased team size can be observed in all patenting subfields (Wuchty et al. 2007). A number of reasons for this development have been put forward. While

Reprinted from Journal of Evolutionary Economics 22(3), 459–480, Springer (2012).

H. Graf (✉)
School of Economics and Business Administration, Friedrich Schiller University Jena, Carl-Zeiss-Str. 3, 07743 Jena, Germany
e-mail: holger.graf@uni-jena.de

G. Buenstorf et al. (eds.), *The Two Sides of Innovation*, Economic Complexity and Evolution, DOI 10.1007/978-3-319-01496-8_4,

some authors suggest that this is due to the increasing capital intensity of research in laboratory sciences (de Solla Price 1963), or an increasing division of labor between specialized scholars (Jones 2005), Wuchty et al. (2007) argue that this is unlikely to explain similar increases in fields with low capital intensity such as the social sciences, or in fields in which the growth of the knowledge stock has been comparably slow. The increasing connectedness of scientists also becomes visible in the changing structure of co-authorship or co-inventor networks, for example, in the formation of giant components, i.e. large groups of interconnected actors (Newman et al. 2002). Guimerà et al. (2005) show how increasing specialization leads to the emergence of such large communities.

However, there are also differences between such networks that are not explained by these general tendencies, but are related to the costs and benefits of forming linkages. Ejermo and Karlsson (2006) argue that, in certain technologies, there is a need to combine pieces knowledge from various fields, while in other technologies, inventions can be developed from a much narrower knowledge base. Accordingly, I suggest that differences in team building and component size distribution are based on the nature of the technology and economic incentives for team building. Innovation—or, more broadly, the process of knowledge generation—is viewed as a directed search process, where the direction of search is guided by technological and economic opportunities. Its path-dependent nature manifests itself in technological trajectories which are guided by paradigms that are only disrupted in few instances (Dosi 1982). Technological opportunities are inherently dynamic and might decline over the course of an industry life cycle as the variety of technological solutions decreases in the paradigmatic stage (Malerba and Orsenigo 1997). However, technologies are not to be seen in isolation. While opportunities in terms of variety decrease over time within a specific field, new opportunities can emerge through combinations of previously unrelated fields, i.e. through a changing *source*, or through an increasing *pervasiveness* when the technology is applied to new markets (Malerba and Orsenigo 1997). For example, early publications by sociologists that tackled economic problems in applying social network analysis opened the door for many economists to follow this approach. Such interdisciplinary tendencies can be observed in many scientific or technological fields. However, in most cases, scientists and engineers are not educated in such an interdisciplinary way. As such, teams of authors of publications or patents become more prevalent than individual authors, especially in fields that connect to other fields. On the basis of these considerations, the central hypothesis is developed which states that technologies which become more central to the knowledge base are also characterized by a higher connectedness of the inventor network.

The theoretical considerations are exemplified in a comparative study of patenting fields over the period 1995 to 2001. In a first step, the technological knowledge base is mapped, resulting in a 'network' of technologies, and changes within its structure are analyzed to identify comparable patenting fields that show distinct characteristics. Key technologies are defined as holding a central position within the knowledge base. It turns out that information technology (IT) shows the highest increase in patenting, while semiconductors shows the highest movement towards the center of the knowledge base, i.e. semiconductors developed, by far, more in the direction of a key technology than did IT.

Since both fields are quite similar in many other dimensions, the evolution of the respective inventor networks is analyzed in a second step. For the construction of inventor networks, I assume that patent inventors are related to their co-inventors as documented on the patent application (Lissoni 2010; Breschi and Catalini 2010). The dynamic analysis of both networks shows an increasing connectedness and the emergence of a large component in semiconductors but not in information technology, which is in line with the expectations.

The remainder of the paper is structured as follows. Section 2 presents a short review of the literature that constitutes the building blocks of my argument. The analysis of the German knowledge base and the identification of interesting objects of analysis is conducted in Section 3. In Section 4, the structural characteristics and changes that occurred in the IT and semiconductor inventor networks are studied. The main results are discussed in the concluding Section 5.

2 Theoretical Building Blocks

2.1 Cumulativeness and Changing Technological Opportunities

In his highly influential article on *the nature of the innovative process*, pp. 222–223 Dosi (1988) presents five stylized facts about innovation that are shared by most, if not all, scholars of innovation: (i) *uncertainty*, which is not simply the lack of all relevant information about the occurrence of known events, but also the existence of techno-economic problems the solution procedures of which are unknown, resulting in the inability to predict precisely the consequences of one's actions; (ii) *increasing reliance* of major new technological opportunities on advances in scientific knowledge; (iii) *increasing complexity* of research and development activities, which causes such activities to be more formally organized rather than carried out by individual innovators; (iv) *increasing role of experimentation* in the form of learning-by-doing and learning-by-using; and (v) the *cumulative character* of innovative activity.

Here, I will concentrate on two of these facts, namely, on the cumulative character of innovative activity and on the sources of technological opportunities. The path-dependent nature of innovation processes manifests itself in trajectories which are guided by technological paradigms (Dosi 1982). Along these trajectories, knowledge is accumulated by generating knowledge built on previous findings, just as Isaac Newton noted in 1676 "If I have seen further it is only by standing on the shoulders of giants."[1] An increasing knowledge stock will then have consequences

[1] The metaphor of a dwarf standing on the shoulder of a giant is actually attributed to Bernard of Chartres and dates back to the twelfth century. See http://en.wikipedia.org/wiki/Standing_on_the_shoulders_of_giants, October 4, 2011.

for later scholars, as they will have to learn more and more about their field before being capable of developing something new. This might lead to either a longer time before researchers can produce new knowledge or it might lead to specialization which demands team building to tackle complex scientific problems.

While the direction of the search process for new ideas is surely dependent on past achievements, it is also guided by actual and perceived opportunities. Opportunities in terms of deepening a technological field are often subject to decreasing returns. At a point where advances can only be made at very high effort and cost, actors will generate new opportunities through combinations of previously unrelated fields, i.e. through a changing *source*, or through an increasing *pervasiveness* when the technology is applied to new markets (Malerba and Orsenigo 1997). In one industry, plenty of opportunities might arise from advances in the sciences; in others, changes in the demand structure might call for new solutions (Pavitt 1984). Of course, even firms within the same industry show different innovation behaviors depending on their strategies (de Jong and Marsili 2006; Leiponen and Drejer 2007) or depending on the technology they apply (Archibugi 2001). For example, firms exploiting an existing technology seem to perform better within a dense cooperation network, while firms with an exploration strategy are better off in a loose network characterized by structural holes (Rowley et al. 2000). Apart from that, these patterns are far from static. Many industries follow a life cycle in which not only the market structure and demand conditions change, but also opportunities for generating novelty. The literature on the industry life cycle provides us with many examples of the way in which innovation shifts from product innovation (generating variety) to process innovation in large scale manufacturing, such as automobiles, penicillin, TV, or tires (Klepper 1996; Buenstorf and Klepper 2009). Interestingly, there are also counterexamples such as the laser industry (Buenstorf 2007), where such a development (at least so far) did not take place. One explanation here is the ongoing formation of submarkets by specialized producers (Klepper and Thompson 2006), which might be considered as a form of increasing pervasiveness.

In both cases, changing source or increasing pervasiveness, a technological field is subject to changes in its relatedness to other fields or in its relative position within knowledge space. A technology that holds a central position in knowledge space is to be considered a key technology that functions as a source of novelty not only in terms of its own development, but for several other fields as well.

2.2 Knowledge Relatedness

The issue of measuring relatedness between different fields is far from new, and various approaches have been taken to broaden our understanding. Scherer (1982) or Pavitt (1984) use information on knowledge flows to measure relatedness between industries and show the way in which knowledge produced in one sector influences progress in others. Patents have been used by Jaffe (1989) to show that the productivity of R&D varies systematically across clusters of technologically related firms, and that this variation is related to the notion of "technological opportunity." Teece et al. (1994) introduce the theoretical concept of technological coherence and derive implications for firms' diversifying strategies that are confirmed by Breschi et al. (2003), who show that firms diversify only into related fields of their existing technology portfolio. Accordingly, Nesta and Saviotti (2005) show that firms with a coherent knowledge base are more successful in innovation. In a related study, Nesta and Saviotti (2006) show the importance of knowledge integration for firms' stock market value in biotechnology.

This bulk of research suggests that knowledge flows or spillovers tend to occur more intensely between related fields, but this literature does not address the issue of the dynamics of knowledge relatedness and how it might be shaped by actors' decisions to direct their search towards new opportunities. Before tackling this problem, let us move into some empirical regularities regarding the organization of the innovation process on a more micro-level, i.e. in network studies of invention and innovation.

2.3 Knowledge Spillovers Through Interpersonal Relationships

There is a long tradition of research on knowledge spillovers. Already in 1890, Alfred Marshall identified informational spillovers as one of the factors leading to the agglomeration of economic activity. He argued that clustered firms have a better production function than isolated firms (see Krugman 1991). In a highly recognized study, Jaffe et al. (1993) show that geographically concentrated patent citations can be interpreted as signs of "localized knowledge spillovers". However, recent attempts to identify channels of knowledge spillovers challenge the argument by Jaffe et al. (1993). Thompson and Fox-Kean (2005), for example, find no intranational localization effects when using a finer level of technological aggregation in their sample. Breschi and Lissoni (2009) argue that there are more and possibly other relevant dimensions than just geographical proximity and introduce a measure of social proximity between inventors to their experiment. They find that social proximity explains most of the identified spillovers and argue that geographical proximity merely facilitates these face-to-face contacts, but geographical proximity is certainly not a sufficient condition for knowledge transmission. These arguments are supported by Singh (2005), who performs a similar study using US patent data.

The importance of labor mobility for knowledge flows is identified by various authors (Zucker et al. 1998; Almeida and Kogut 1999; Møen 2005). As such, interpersonal networks are considered an important channel for the diffusion of knowledge and information (e.g. Zander and Kogut 1995; Sorenson 2003). In the same spirit, Ejermo and Karlsson (2006) argue that such direct interaction should be more substantial than citations as indicators of the overall flows of knowledge within an innovation system. The consequence of these findings is that, to benefit from knowledge spillovers, actors have to establish relations to others and thereby position themselves within a social network.

2.4 Inventor Networks: Increasing Team Size and the Formation of Large Components

Within the literature on invention and innovation, we find various types of networks. Depending on the research question, nodes might be firms, individuals, regions, patents, etc. Linkages can be any type of relation between these nodes, such as teacher-student relationship, citation, informal communication or collaboration resulting in jointly authored publications (Crane 1969; Andersson and Beckmann 2009). Given the importance of interpersonal networks for knowledge transmission, I focus on networks with individuals as nodes.[2] Individuals are commonly linked through documented joint work. In the case of co-authorship networks, the information is based on publication records, and two authors are linked if they have jointly written a paper. An inventor network is a special type of co-authorship network where linkages are established through joint patents instead of publications. The underlying assumption is that co-authors or co-inventors know each other and have exchanged some information or even learned from each other (Lissoni 2010).

In scientometrics, there is a long tradition of exploiting information on co-authorship of scientific publications to analyze knowledge exchange among researchers and to investigate social networks of academic scientists (Persson and Beckmann 1995; Melin and Persson 1996). Recent contributions are especially interested in the structural properties of such networks in the search for generic mechanisms driving their development (Barabasi et al. 2002; Newman 2001; Newman et al. 2002; Wagner and Leydesdorff 2005). However, there are also structural differences between scientific fields (Barabasi et al. 2002) or even within a narrowly defined area (Moody 2004). The structure of networks is also said to influence system performance. Fleming et al. (2007) show that differences in regional innovative performance can be traced back to the connectedness of the respective inventor networks. Cowan and Jonard (2003, 2004, 2007a, b) study the way in which different network structures affect the flow of knowledge by means of simulation.

Over the second half of the 20th century, the generation of knowledge seems to have been subject to a major shift towards an increasing dominance of teamwork (Wuchty et al. 2007). This development can be observed across all fields of research. In the sciences, average team size almost doubled between 1955 and

[2] Ejermo and Karlsson (2006) provide a very detailed account of the motives and benefits of forming such networks and discuss implications for their evolution.

2000; while, in 1956, 17.5% of all papers were written by teams, this figure increased to 51.5% in 2000. In patenting, average team size increased from 1.7 to 2.3 inventors per patent between 1975 and 2000, and this tendency towards increased team size is observed across all patenting subfields (Wuchty et al. 2007). While some authors suggest that this is due to the increasing capital intensity of research in laboratory sciences (de Solla Price 1963), this cannot explain why this sharp increase in teamwork is also observed in the social sciences or economics. Another explanation is an increasing division of labor between specialized scholars (Jones 2005). However, Wuchty et al. (2007) argue that this is unlikely to explain similar increases in fields in which the growth of the knowledge stock has been comparably slow. Others argue that the costs of communication have decreased, making collaboration between researchers from different places in the world more attractive.

The way teams are formed and how this extends to the formation of networks have attracted considerable research. Newman et al. (2002) state that real world social networks without a giant component rarely exist. In their random graph model, they derive properties of the degree distribution that predict the formation and size of such a component. In a longitudinal study on network formation of all patenting inventors in the US from 1975 to 2002, Fleming and Frenken (2007) observe a sudden rise of a giant component in Silicon Valley but not in the Boston region. Their result is explained by differences in the rate of labor mobility between the two regions, which certainly affects the degree distribution. Guimerà et al. (2005) investigate the manner in which the mechanisms by which creative teams self-assemble and determine the structure of these collaboration networks. They propose a model for the self-assembly of creative teams that has its basis in three parameters: team size, the fraction of newcomers in new productions, and the tendency of incumbents to repeat previous collaborations. In their model, increasing specialization leads to the emergence of a large connected community of actors.

2.5 Hypothesis

The previous discussion suggests a mechanism that could be responsible for differences in team building and component size distribution when comparing inventor networks across technologies. The argument, which is developed below, is complementary to existing explanations of general trends in inventor networks. To recapitulate, the process of knowledge generation is viewed as a directed search process, where the direction of search is guided by technological and economic opportunities. Technological opportunities might decline over the course of an industry life cycle if the variety of technological solutions decreases in the paradigmatic stage (Malerba and Orsenigo 1997). However, while opportunities within a specific field might diminish over time, new opportunities can emerge through new combinations of previously unrelated fields, i.e. through a changing *source*, or through an increasing *pervasiveness* when new applications are identified for a technology (Malerba and Orsenigo 1997).

The direction of the search process should influence the position of a technology within the knowledge base. In cases where opportunities are to be found in knowledge deepening activities, we should expect to observe a relatively stable position within the knowledge base. An example could be research on the fuel efficiency of internal combustion engines. The existing technology might be refined through a better arrangement of the mechanical parts to optimize efficiency, but there is no need to connect to different technologies in the development, nor does the better engine apply to other technologies. If opportunities arise from widening activities through combinations of previously unrelated fields, the technology is expected to move towards the center of the knowledge base and to develop characteristics of a key technology. To pick up on the above example, fuel efficiency might also be enhanced by having an electronic engine control unit determine the amount of fuel, ignition timing and other parameters to increase efficiency. Electronics is thereby connected to mechanical engineering to generate novelty through new combinations; mechanical engineering finds a new source and electronics becomes more pervasive.

Such interdisciplinary tendencies can be observed in many scientific or technological fields. However, in most cases, scientists and engineers are not educated in such an interdisciplinary way. As such, teams of authors of publications or patents become more prevalent than individual authors, especially in fields that connect to other fields. We can easily imagine that it is possible for a single capable engineer to increase fuel efficiency by merely working on the mechanical parts, but for the development of an engine control unit, we might need a mechanical engineer, an electronic engineer, and possibly a computer scientist for the software. My argument is not much different from the explanation of Jones (2005), but adds a deeper understanding of the reason why we observe an increasing formation of interdisciplinary teams and why this process differs between fields of research. On the basis of these considerations, the central hypothesis states that technologies which become more central within the knowledge base are characterized by a higher connectedness of the respective inventor network.

This connectedness might be measured in two ways. On the level of the individual inventor, we should observe an increasing degree centrality (number of distinct collaborators) within technological fields that move towards the center of the knowledge base. Regarding the overall structure of the inventor network, we would expect the formation of a large component in such a technology.

3 Mapping the Technological Knowledge Base

The above theorizing leads to a research design that has to be performed in two steps. In a first step, the knowledge base of Germany is constructed to identify technologies that show significant developments in terms of patenting or in terms of their position within the knowledge base. In a second step, these interesting cases are then analyzed in terms of their inventor networks in Section 4.

In this setting, the knowledge base is to be understood as a network of interrelated parts, where two fields of knowledge are related if new knowledge (in this case a patent, but this exercise could be performed with academic papers as well) touches both of these fields.[3] The more often two fields are connected, the closer they are assumed to be in technology space (Nesta and Saviotti 2005, 2006).

Applications at the German patent office are used to map the knowledge base for each year from 1995 to 2001. The publishing date is used to avoid truncation. On each patent, several IPC classes are named and they cover the technological areas for which the patent is relevant. These classes are either provided by the applicant or by patent examiners during the process of review and specify the relevant technological fields. Since patents that share common technological classes, are assumed to be more closely related than other patents, the frequency of co-occurrence is assumed to be proportional to the intensity of the linkage (Saviotti 2007). This information is employed to link the n technologies in the network in the following way: if a patent is in class i and in class j, there is one link between nodes i and j. The more co-occurences of two technologies, J_{ij} are observed, the stronger these two are assumed to be related. Notice that resulting relations between technological classes are not directed as they would be if citation data were used.

Following the notation of Nesta and Saviotti (2006), Ω is the resulting $n \times n$ matrix of co-occurences, which can be thought of absolute technological relations:

$$\mathbf{\Omega} = \begin{pmatrix} J_{11} & \cdots & J_{1j} & \cdots & J_{1n} \\ \vdots & \ddots & & & \vdots \\ J_{i1} & & J_{ij} & & J_{in} \\ \vdots & & & \ddots & \vdots \\ J_{n1} & \cdots & J_{nj} & \cdots & J_{nn} \end{pmatrix} \tag{1}$$

However, technological classes in which many patents are assigned, have a higher chance of co-occurrence. Therefore, a measure of relatedness is constructed based on the matrix Ω as in Nesta and Saviotti (2006). For the calculation of the expected number of co-occurences, a random process is assumed in which the probability of a patent being assigned to technologies i and j is a hypergeometric random variable with a mean of μ_{ij} and a variance of σ^2_{ij}:

$$\mu_{ij} = \frac{O_i O_j}{K}, \tag{2}$$

[3] I acknowledge that a patent is a measure of invention rather than innovation, but since the argument is based on innovation as a search process, the implementation of an invention within an innovation is not necessary for a patent to serve as a proxy for new knowledge.

$$\sigma_{ij}^2 = \mu_{ij}\left(\frac{K - O_i}{K}\right)\left(\frac{K - O_j}{K - 1}\right), \tag{3}$$

where K is the total number of patents, O_i of which are assigned to technology i and O_j in technology j. Relatedness is then defined as

$$\tau_{ij} = \frac{J_{ij} - \mu_{ij}}{\sigma_{ij}}, \tag{4}$$

which is positive for technologies that co-occur more often than expected ($J_{ij} > \mu_{ij}$) and negative if $J_{ij} < \mu_{ij}$.

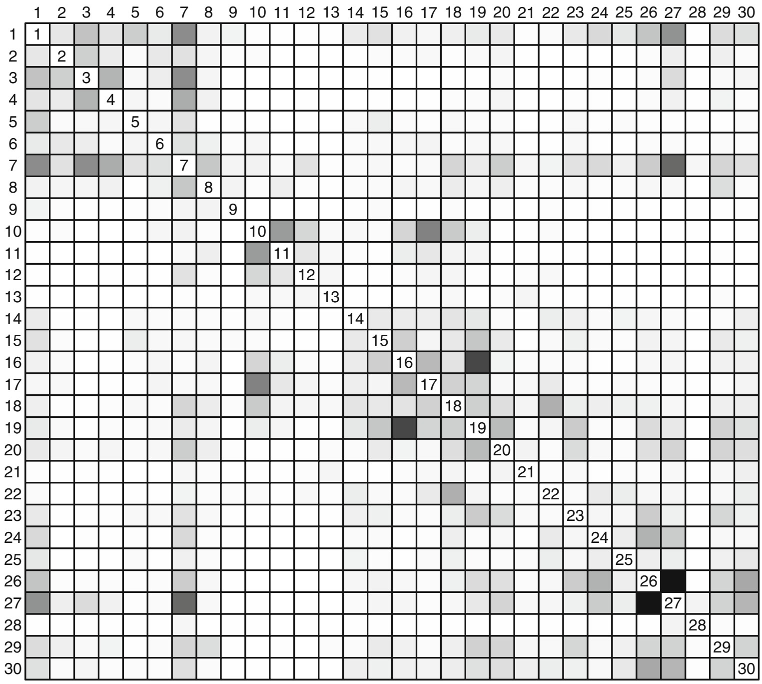
For the technological aggregation of IPC classes, patents are classified according to a technology-oriented classification that distinguishes 5 sectors and 30 technologies based on the International Patent classification (IPC). This classification has been elaborated jointly by the 'Fraunhofer-Institut 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) and is provided in Table 2 in Appendix.

In Fig. 1, the German knowledge base is visualized for the whole period (1995 to 2001) in the form of a levelplot. Figure 1a displays the network of co-occurences, while, in Fig. 1b, the relatedness between technologies τ_{ij} determines the coloring of cells.

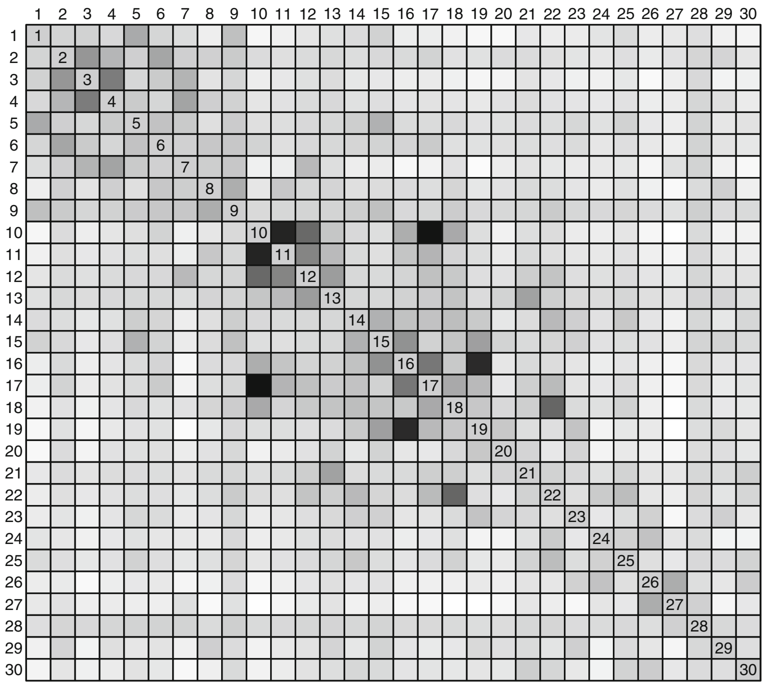
Strong connections can be observed between technologies 1 to 7 which covers five technologies within the sector of electrical engineering, together with optics and analysis, measurement and control technology. Another group of connected fields is in technologies within large scale chemistry and process engineering (14 to 19). What can also be read from this plot are the generic features of analysis, measurement and control technology (7) and electrical machinery (1), which show high levels of connection to many different fields. Especially strong linkages exist between 16 (Macromolecular chemistry, polymers) and 19 (Materials processing, textiles, paper), and between 26 (Mechanical elements) and 27 (Transport).

While the general structure of both figures is the same, it is noteworthy that, in Fig. 1b, high patenting fields such as transport (27) or analysis, measurement and control technology (7) are not as much related to other fields as if the absolute number were considered. Since it is controlled for the absolute number of patents in Fig. 1b, these differences are not surprising, but provide a rationale to analyze the knowledge base according to co-occurences in the following.

This analysis provides a static picture of the German knowledge base covering patenting activities over a seven year period. It represents the strengths of the German innovation system in automobiles and chemicals as rather distinct subsystems with instruments connecting the knowledge base. However, the primary interest lies in the changes that occurred during that period. The late 1990s were characterized by major advances in fields such as IT, semiconductors, and biotechnology, and it should be interesting to observe the manners in which these



(a) Co-occurences of technologies i and j (Jij)



(b) Relatedness of technologies i and j (τij)

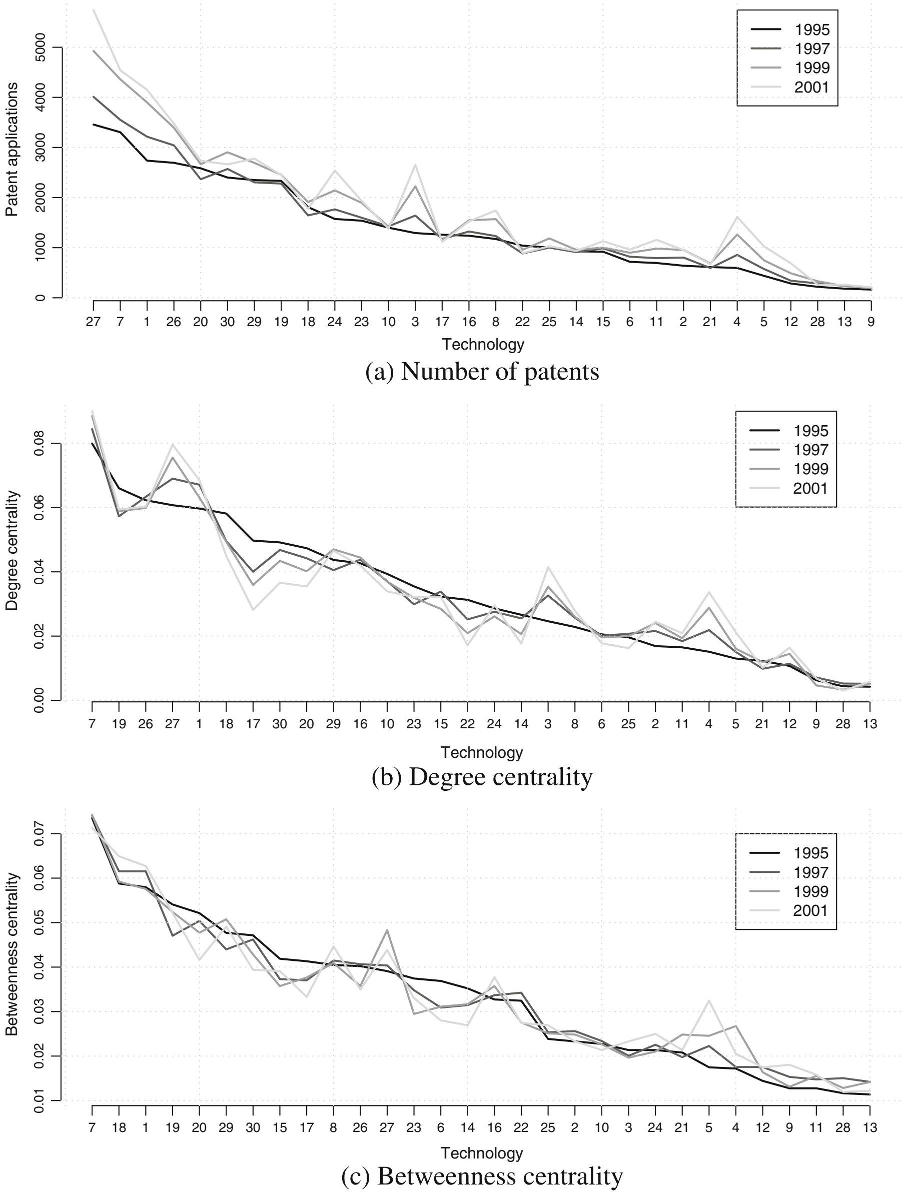
Fig. 1 Knowledge base of Germany 1995–2001 advancements changed the technological knowledge base. For that purpose, changes in the importance of technologies are displayed according to three different variables in Fig. 2. Overall, correlations of these variables between different years suggest quite a stable structure (they range between 0.89 and 0.99), but we can observe significant changes for some fields.

In Fig. 2a, technologies are arranged in descending order according to the number of patents in 1995 (the black line). Values for subsequent periods (1997, 1999, 2001) are plotted as lines of decreasing darkness by keeping the same order. Even though there is no direct connection between two neighboring fields, the lines help us to visualize changes within the knowledge base. We observe a deepening of the knowledge base in the sense that strong patenting fields such as transport (27), analysis (7), and electrical machinery (1) all show increasing numbers of patent applications. Some fields become decidedly stronger, such as engines (24), telecommunications (3), and information technology (4). While these developments in patenting indicate performance within different fields, they do not tell us the way in which these technologies position themselves within the knowledge base.

Since the matrix of co-occurrences constitutes a network, I propose that the position of a technology can be measured by graph theoretic concepts. Degree centrality counts the linkages of one technology to all others, where the linkages are weighted by the number of co-occurrences. This is simply the sum across all columns of Ω. Figure 2b visualizes changes in the degree centrality of technologies, i.e. how strong a technology is connected to all others. Here we see that analysis (7) remains the most central technology, and transport (27) is becoming more connected, while a sharp decrease in connections is observed for the chemical industry (17). Degree has one important shortcoming if we want to apply it to the concept of a key technology. It does not account for the variety of connections of a technology to other fields. The high values in transport, for example, arise because of its strong relations to mechanical elements but not to the multitude of other fields. As such, I consider betweenness centrality to be a better indicator for key technologies. In social network analysis, betweenness centrality is measured as the frequency with which a node is positioned between pairs of other nodes on the shortest path connecting them.[4] This measure comes closer to our perception of a key technology, since it accounts for the variety of sources and applications of a technology during the inventive phase.

For the knowledge base of Germany, we find analysis (7), chemical engineering (18), and electrical machinery (1) to be the three most important and persistent key technologies. The most drastic increase in this measure appears for semiconductors (5), where betweenness more than doubles. This indicates that the sources and applications of inventions in the field of semiconductors became more diverse during this period. New applications are, for example, found in patents that connect semiconductors with biotechnology, a combination that was not apparent in 1995.

[4] The values in the co-occurrence matrix are interpreted as capacities of the linkages.



(a) Number of patents

(b) Degree centrality

(c) Betweenness centrality

Fig. 2 Changes in the importance of technological classes for the German knowledge base

The number of patents that combine semiconductors and engines in 2001 is six times as high as in 1995. We also observe sharp increases in the connection to handling in printing, which indicates changes in the production processes of semiconductors.

To provide a more structured impression of the development of technological positions within the knowledge base, growth rates of the number of patents and betweenness centrality are calculated for the period 1995 to 2001. In Fig. 3, where these growth rates are plotted against each other, two interesting cases appear. IT (4) and semiconductors (5) seem to follow quite distinct patterns in terms of integration with other technological fields. While IT shows high increases in patents, betweenness centrality only increases slightly. By contrast, semiconductors shows a very high growth rate in betweenness centrality, while the increase in patenting is less pronounced.

4 Inventor Networks in IT and Semiconductors

One question that arises from the exercise above, concerns the relation between i) the integration of a specific technology—or its development—within the knowledge base and ii) the organization of the innovation process within a specific technology. One can imagine that the organization of the innovation process differs between technologies that deepen their specific knowledge base and those that widen its knowledge base in establishing connections to other fields. The development of patents that link different technological fields clearly needs different capabilities than the development of patents that only touch one field. Apparently, competencies for widening knowledge might not be found within a single person, but we should expect them to arise from the collaboration within teams of inventors. As such, a field that is increasingly central within the knowledge base should also be characterized by increasing interaction between inventors.

To provide an answer to this question, an analysis of the inventor networks is performed for the two fields identified in the previous section: IT (4) and semiconductors (5). Specifically, from the original data base, all patents with an IPC code (main or supplementary class) that falls into these two technologies are extracted. For the purpose of comparison, it is fortunate that these two fields belong to the same sector and are also quite similar in terms of the number of patents and inventors. Furthermore, the top applicants show up in both technologies. Siemens, Bosch, and Infineon are among the top five, while Philips, Daimler and Fraunhofer are among the top 10 applicants in both fields. This is important since it makes an interpretation on technological grounds easier if it can—at least partly—be excluded that different firms strategies or governance mechanisms shape these two networks.

To reconstruct inventor networks, inventors are related through common team membership as documented on patents; i.e. two inventors are related if their collaboration is documented by their common naming on the same patent (Balconi et al. 2004; Fleming et al. 2007). Since knowledge exchange between individuals does not occur at one point in time, the common procedure in constructing such networks is to consider patents from more than one year. I decided to take three-year

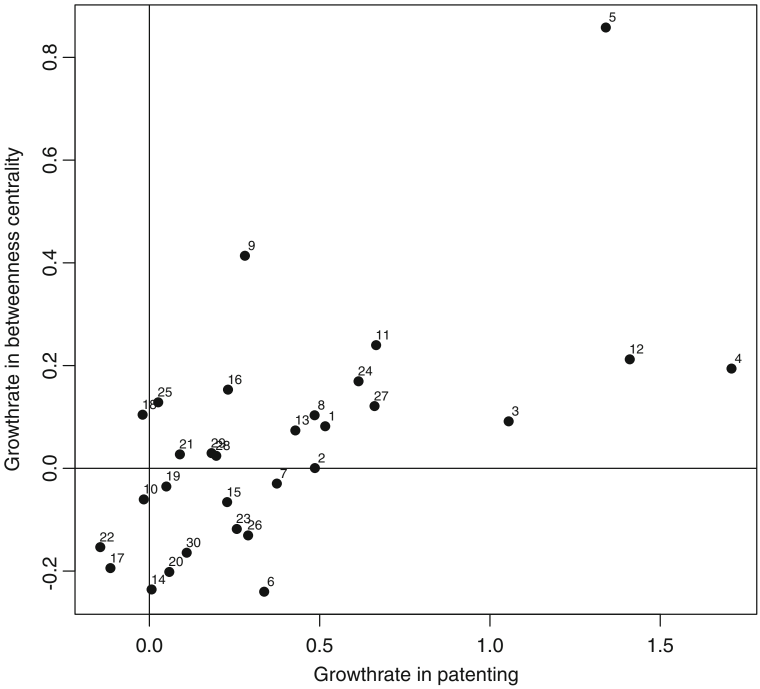


Fig. 3 Scatterplot of growth rates in patenting and betweenness centrality of technological classes

moving windows,[5] which implies fewer connections than in longer periods but provides more observations to analyze the development of the networks. This also implies that relationships between team members are broken up if they are not renewed in terms of collaboration.

Table 1 presents the descriptive statistics of these two networks for the years 1997 to 2001 to analyze gradual changes in the network structure. A comparison of the development of the inventor networks in these two technologies shows some interesting features. Information technology is the larger network in terms of inventors and also grew faster between 1997 and 2001. The growth rate of the IT network between the first and last year is 0.795, while the semiconductor network grew by 0.464. Since information technology was chosen because of its sharp increase in patenting, this result is not surprising. However, the connectedness of the two networks shows strong differences. One indicator of the connectedness of a network is the share of actors within the main component. Components are connected parts of the network and the main component is simply the one with the highest number of members. The formation of large components has been analyzed by Fleming and Frenken (2007) for regional networks. In IT, we observe no such formation with the number of inventors starting at 28 in 1997, peaking at

[5] Fleming et al. (2007) e.g. use a five-year moving window.

Table 1 Dynamics of the inventor network in IT and semiconductors

	1997	1998	1999	2000	2001
Information technology (4)					
Nodes	2,946	3,430	4,146	4,584	5,287
Number of components	1,460	1,716	2,067	2,277	2,533
Size of main component	28	107	87	53	59
Share in main component	0.95%	3.12%	2.10%	1.16%	1.12%
Isolates	784	952	1138	1272	1379
Share of isolates	26.61%	27.76%	27.45%	27.75%	26.08%
Centralization	0.0102	0.0195	0.0140	0.0098	0.0109
Density	0.0007	0.0006	0.0005	0.0004	0.0004
Mean degree	1.9579	2.0093	2.0014	2.0414	2.1865
Semiconductors (5)					
Nodes	2,122	2,394	2,694	2,849	3,107
Number of components	754	813	887	908	934
Size of main component	78	256	291	334	475
Share in main component	3.68%	10.69%	10.80%	11.72%	15.29%
Isolates	307	341	363	397	404
Share of isolates	14.47%	14.24%	13.47%	13.93%	13.00%
Centralization	0.0199	0.0242	0.0200	0.0220	0.0179
Density	0.0014	0.0013	0.0012	0.0012	0.0011
Mean degree	2.9161	3.1997	3.2858	3.5114	3.5558

Networks are reconstructed from patent applications within a three-year window, up to the specified year

107 in the following year, and declining to a value of 59 in 2001. Given the size of this network, this is a very small main component and, accordingly, we observe a share in the range between 1% and 3% of all inventors within that main component. This volatility implies a high fragility of this largest connected part of the network. The picture of semiconductors is a different one. Here, the share of inventors in the main component starts at 3.7% and continually rises up to 15.3% in 2001. This result meets the expectation of an increased connectedness in a technology that becomes more central within the knowledge base.

If we consider interdisciplinary teams to be a reaction to technological opportunities that arise from an increased pervasiveness, isolated inventors who have no contacts to other actors should also be less common in a widening technology compared to a deepening one. Again, in line with the expectations, the share of isolated inventors in IT is almost twice as high compared to semiconductors and it seems to be much easier to generate novelty as a single inventor in IT than it is in semiconductors.

Centralization is not so much a measure of connectedness than of the concentration of inventive activities among inventors. In Table 1, the centralization measure based on degree centrality, i.e. on the number of linkages of an actor, is reported. The theoretical minimum of this index is 0 for a ring structure or any network in which every actor has the same number of linkages, and its maximum is 1 for networks in which all actors are connected to a single, central actor (star network). In both networks, centralization takes rather low values between 0.01 and 0.025 and is always higher in the semiconductor network.

The network density is defined as the number of all linkages divided by the number of possible linkages and gives an impression of the connectedness of the network. This measure is somewhat problematic in comparing networks of different sizes, as the number of possible linkages increases geometrically, while the actual number of linkages usually does not, since inventors are constrained in their capacities to have contacts to other actors. Therefore, the mean degree, i.e. the average number of ties, is also reported. In both networks, the mean degree increases steadily, indicating a general tendency towards an increasing relevance of teamwork, which is in line with the results by Wuchty et al. (2007). But again, we observe sharp differences between both technologies. While in IT the mean degree is comparably low and increases only slightly from 1.96 in 1997 to 2.18 in 2001, values in semiconductors increase twice as fast and at a higher level from 2.92 to 3.56 in semiconductors.

The central statistics for a comparison of the connectedness in the inventor networks of IT and semiconductors are presented in Fig. 4 to illustrate the sharp differences in the share of inventors within the main component and the mean degree.

5 Conclusions

Many scholars have performed research on the way in which sectors, industries, and technologies are interrelated and how knowledge flows from one field to the other, or how certain fields draw on the knowledge produced in other fields. Quite often this research comes with an explicit or implicit criticism on existing classifications of industry, but few consider these relations to be of a changing nature. An exploratory exercise was conducted to get an impression of how such a process could be explained and the consequences it might have for the organization of inventive activities. More specifically, I analyzed the relationship between the changing position of technologies within the knowledge base and the changes within the underlying inventor networks, i.e. the structure of collaborating actors who produce new knowledge and thereby not only constitute but also actively change the knowledge base and react to new opportunities.

In addition to existing explanations of increasing team size and the formation of large components, I could identify another mechanism in which the nature of technological opportunities plays an important role for the incentives to form larger teams in the process of invention. If opportunities within a specific domain are still abundant, novelty is generated by deepening existing knowledge and collaboration with a diverse set of actors plays only a minor role. Whenever opportunities arise from broadening the field of applications or from drawing on the knowledge of a variety of technologies, interdisciplinary teams are formed which constitute network structures of higher connectivity among inventors. I propose that the nature of

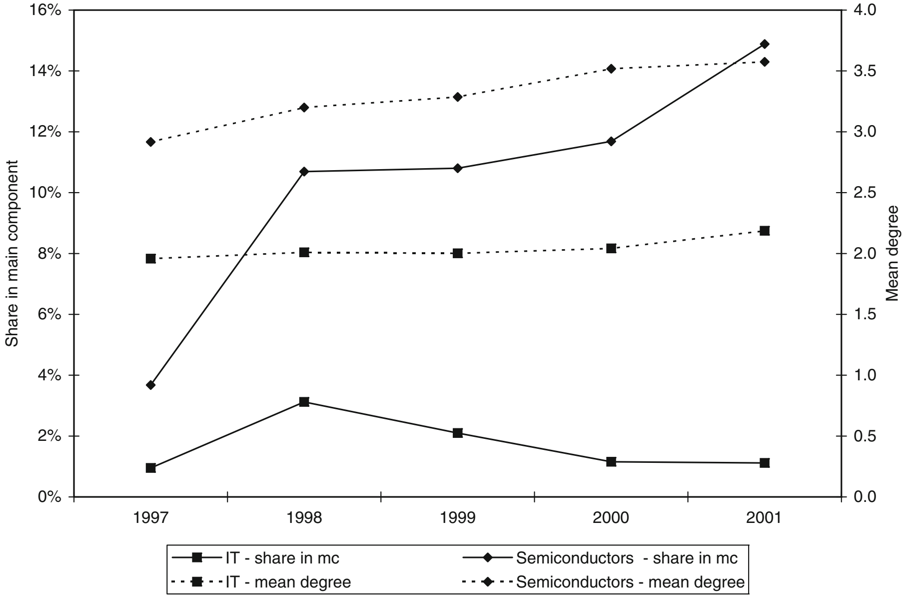


Fig. 4 Connectedness of inventor networks in IT and semiconductors

technological opportunities is driving the process that determines whether a certain field moves towards the center of the knowledge base.

This economic argument for differences in the intensity of collaboration between technological fields has non negligible implications for policy. Several policy programs aim at bringing together actors irrespective of the characteristics of the technology. In fields of activity in which opportunities are to be found in deepening knowledge, such a policy might induce excessive networking. As such, knowledge about the characteristics of opportunities might increase the precision of policy measures.

This research has to be viewed as a first step into the analysis of a co-evolution of technology and invention networks and certainly has its limitations. The direction of causality is assumed to run from the identification of new opportunities to the formation of interdisciplinary teams. In principle, however, causality could also run in the opposite direction. It might well be that through experimentally setting up interdisciplinary teams, their success opens up these new opportunities. From a methodological point of view, I have to acknowledge that patents as a data source have their flaws which might well affect the results. Many innovations are not patented, which is the reason why we can only observe the tip of the iceberg of actual relations in the innovation process. The fact that patents are increasingly used for strategic purposes (Cohen et al. 2000; Blind et al. 2006, 2009) might also mislead us, since no teams of inventors are needed for patents that do not reflect actual progress in knowledge production. For the analysis of such a co-evolutionary process, the time span covered is certainly short and similar studies will have to test these results with developments within a broader set of technological areas.

Acknowledgements During the process of conducting this research, I benefited from various discussions with Uwe Cantner. I would also like to thank Alexander Ebner, Michael Fritsch, Andreas Pyka for helpful comments on earlier drafts. Comments and suggestions by two anonymous referees are greatly acknowledged. This paper was presented at the Jahrestagung des evolutorischen Ausschusses des Vereins für Socialpolitik 2009 in Jena, at the EAEPE Conference 2009 in Amsterdam and at the Conference of the International Schumpeter Society 2010 in Aalborg. Of course, I am responsible for any remaining errors or shortcomings.

Appendix

Table 2 Concordance between IPC and technology codes

Industry technology	IPC-code
I. Electrical engineering	
1. Electrical machinery and apparatus, electrical energy	F21; G05F; H01B, C, F, G, H, J, K, M, R, T; H02; H05B, C, F, K
2. Audiovisual technology	G09F, G; G11B; H03F, G, J; H04N-003, -005, -009, -013, -015, -017, R, S
3. Telecommunications	G08C; H01P, Q; H03B, C, D, H, K, L, M; H04B, H, J, K, L, M, N-001, -007, -011, Q
4. Information technology	G06; G11C; G10L
5. Semiconductors	H01L, B81
II. Instruments	
6. Optics	G02; G03B, C, D, F, G, H; H01S
7. Analysis, measurement, control technology	G01B, C, D, F, G, H, J, K, L, M, N, P, R, S, V, W; G04; G05B, D; G07; G08B, G; G09B, C, D; G12
8. Medical technology	A61B, C, D, F, G, H, J, L, M, N
9. Nuclear engineering	G01T; G21; H05G, H
III. Chemistry, pharmaceuticals	
10. Organic fine chemistry	C07C, D, F, H, J, K
11. Pharmaceuticals, cosmetics	A61K, P
12. Biotechnology	C07G; C12M, N, P, Q, R, S
13. Agriculture, food chemistry	A01H; A21D; A23B, C, D, F, G, J, K, L; C12C, F, G, H, J; C13D, F, J, K
14. Materials, metallurgy	C01; C03C; C04; C21; C22; B22; B82
15. Surface technology, coating	B05C, D; B32; C23; C25; C30
16. Macromolecular chemistry, polymers	C08B, F, G, H, K, L; C09D, J
17. Chemical industry and petrol industry, basic materials chemistry	A01N; C05; C07B; C08C; C09B, C, F, G, H, K; C10B, C, F, G, H, J, K, L, M; C11B, C, D
IV. Process engineering, special equipment	
18. Chemical engineering	B01B, D (without -046 to -053), F, J, L; B02C; B03; B04; B05B; B06; B07; B08; F25J; F26
19. Materials processing, textiles, paper	A41H; A43D; A46D; B28; B29; B31; C03B; C08J; C14; D01; D02; D03; D04B, C, G, H; D05; D06B, C, G, H, J, L, M, P, Q; D21

(continued)

Table 2 (continued)

Industry technology	IPC-code
20. Handling, printing	B25J; B41; B65B, C, D, F, G, H; B66; B67
21. Agricultural and food machinery and apparatus	A01B, C, D, F, G, J, K, L, M; A21B, C; A22; A23N, P; B02B; C12L; C13C, G, H
22. Environmental technology	A62D; B01D-046 to -053; B09; C02; F01N; F23G, J
V. Mechanical engineering, machinery	
23. Machine tools	B21; B23; B24; B26D, F; B27; B30
24. Engines, pumps, turbines	F01B, C, D, K, L, M, P; F02; F03; F04; F23R
25. Thermal processes and apparatus	F22; F23B, C, D, H, K, L, M, N, Q; F24; F25B, C; F27; F28
26. Mechanical elements	F15; F16; F17; G05G
27. Transport	B60; B61; B62; B63B, C, H, J; B64B, C, D, F
28. Space technology, weapons	B63G; B64G; C06; F41; F42
29. Consumer goods and equipment	A24; A41B, C, D, F, G; A42; A43B, C; A44; A45; A46B; A47; A62B, C; A63; B25B, C, D, F, G, H; B26B; B42; B43; B44; B68; D04D; D06F,N ; D07; F25D; G10B, C, D, F, G, H, K
30. Civil engineering, building, mining	E01; E02; E03; E04; E05; E06; E21

Source: ISI OST INPI classification (update 2005) (Schmoch 2008)

References

Almeida P, Kogut B (1999) Localization of knowledge and the mobility of engineers in regional networks. Manage Sci 45(7):905–917

Andersson ÅE, Beckmann MJ (2009) Economics of knowledge: theory, models and measurement. Edward Elgar, Cheltenham, UK

Archibugi D (2001) Pavitt's taxonomy sixteen years on: a review article. Econ Innov New Technol 10(5):415–425

Balconi M, Breschi S, Lissoni F (2004) Networks of inventors and the role of academia: an exploration of Italian patent data. Res Policy 33:127–145

Barabasi A, Jeong H, Neda Z, Ravasz E, Schubert A, Vicsek T (2002) Evolution of the social network of scientific collaborations. Physica A 311(3–4):590–614

Blind K, Edler J, Frietsch R, Schmoch U (2006) Motives to patent: empirical evidence from Germany. Res Policy 35(5):655–672

Blind K, Cremers K, Mueller E (2009) The influence of strategic patenting on companies' patent portfolios. Res Policy 38(2):428–436

Breschi S, Lissoni F (2009) Mobility of skilled workers and co-invention networks: an anatomy of localized knowledge flows. J Econ Geogr 9(4):439–468

Breschi S, Catalini C (2010) Tracing the links between science and technology: an exploratory analysis of scientists' and inventors' networks. Res Policy 39(1):14–26

Breschi S, Lissoni F, Malerba F (2003) Knowledge-relatedness in firm technological diversification. Res Policy 32(1):69–87

Buenstorf G (2007) Evolution on the shoulders of giants: entrepreneurship and firm survival in the German laser industry. Rev Ind Organ 30(3):179–202

Buenstorf G, Klepper S (2009) Heritage and agglomeration: the Akron tyre cluster revisited. Econ J 119(537):705–733

Cohen WM, Nelson RR, Walsh JP (2000) Protecting their intellectual assets: appropriability conditions and why U.S. manufacturing firms patent (or not). Tech. rep., NBER Working Paper No. W7552
Cowan R, Jonard N (2003) The dynamics of collective invention. J Econ Behav Organ 52(4): 513–532
Cowan R, Jonard N (2004) Network structure and the diffusion of knowledge. J Econ Dyn Control 28(8):1557–1575
Cowan R, Jonard N (2007a) Merit, approbation and the evolution of social structure. J Econ Behav Organ 64(3–4):295–315
Cowan R, Jonard N (2007b) Structural holes, innovation and the distribution of ideas. J Econ Int Coordinat 2(2):93–110
Crane D (1969) Social structure in a group of scientists: a test of the "invisible college" hypothesis. Am Sociol Rev 34(3):335–352
Dosi G (1982) Technological paradigms and technological trajectories: a suggested interpretation of the determinants and directions of technical change. Res Policy 11(3):147–162
Dosi G (1988) The nature of the innovative process. In: Dosi G, Freeman C, Nelson R, Silverberg G, Soete L (eds) Technical change and economic theory, Pinter, London, pp 221–238
Ejermo O, Karlsson C (2006) Interregional inventor networks as studied by patent coinventorships. Res Policy 35(3):412–430
Fleming L, Frenken K (2007) The evolution of inventor networks in the Silicon Valley and Boston regions. Adv Complex Systems 10(1):53–71
Fleming L, King I Charles, Juda AI (2007) Small worlds and regional innovation. Organ Sci 18(6): 938–954
Guimerà R, Uzzi B, Spiro J, Amaral LAN (2005) Team assembly mechanisms determine collaboration network structure and team performance. Science 308(5722):697–702
Jaffe AB (1989) Characterizing the "technological position" of firms, with application to quantifying technological opportunity and research spillovers. Res Policy 18(2):87–97
Jaffe AB, Trajtenberg M, Henderson R (1993) Geographic localization of knowledge spillovers as evidenced by patent citations. Q J Econ 108(3):577–598
Jones BF (2005) The burden of knowledge and the 'death of the renaissance man': is innovation getting harder? Working Paper 11360, National Bureau of Economic Research. http://www.nber.org/papers/w11360
de Jong JP, Marsili O (2006) The fruit flies of innovations: a taxonomy of innovative small firms. Res Policy 35(2):213–229
Klepper S (1996) Entry, exit, growth, and innovation over the product life cycle. Am Econ Rev 86 (3):562–583
Klepper S, Thompson P (2006) Submarkets and the evolution of market structure. RAND J Econ 37(4):861–886
Krugman P (1991) Increasing returns and economic geography. J Polit Econ 99:483–499
Leiponen A, Drejer I (2007) What exactly are technological regimes?: intra-industry heterogeneity in the organization of innovation activities. Res Policy 36(8):1221–1238
Lissoni F (2010) Academic inventors as brokers. Res Policy 39(7):843–857
Malerba F, Orsenigo L (1997) Technological regimes and sectoral patterns of innovative activities. Ind Corp Change 6(1):83–117
Marshall A (1890) Principles of economics. Macmillan, London
Melin G, Persson O (1996) Studying research collaboration using co-authorships. Scientometrics 36(3):363–377
Møen J (2005) Is mobility of technical personnel a source of R&D spillovers? J Labor Econ 23(1): 81–114
Moody J (2004) The structure of a social science collaboration network. Am Sociol Rev 69: 213–238
Nesta L, Saviotti PP (2005) Coherence of the knowledge base and the firm's innovative performance: evidence from the U.S. pharmaceutical industry. J Ind Econ 53(1):123–142

Nesta L, Saviotti PP (2006) Firm knowledge and market value in biotechnology. Ind Corp Change 15(4):625–652
Newman MEJ (2001) Scientific collaboration networks. I. Network construction and fundamental results. Phys Rev E 64(1):016,131
Newman M, Watts D, Strogatz S (2002) Random graph models of social networks. Proc Natl Acad Sci USA 99:2566–2572
Pavitt K (1984) Sectoral patterns of technical change: towards a taxonomy and a theory. Res Policy 13:343–375
Persson O, Beckmann M (1995) Locating the network of interacting authors in scientific specialties. Scientometrics 33(3):351–366
Rowley T, Behrens D, Krackhardt D (2000) Redundant governance structures: an analysis of structural and relational embeddedness in the steel and semiconductor industries. Strateg Manage J 21(3):369–386
Saviotti PP (2007) On the dynamics of generation and utilisation of knowledge: the local character of knowledge. Struct Change Econ Dynam 18(4):387–408
Scherer FM (1982) Inter-industry technology flows and productivity growth. Rev Econ Stat 64: 627–634
Schmoch U (2008) Concept of a technology classification for country comparisons. Final report to the world intellectual property organisation (wipo), WIPO
Singh J (2005) Collaborative networks as determinants of knowledge diffusion patterns. Manage Sci 51(5):756–770
de Solla Price DJ (1963) Little science, big science. Columbia Univ. Press, New York
Sorenson O (2003) Social networks and industrial geography. J Evol Econ 13:513–527
Teece D, Rumelt R, Dosi G, Winter S (1994) Understanding corporate coherence: theory and evidence. J Econ Behav Organ 23:361–377
Thompson P, Fox-Kean M (2005) Patent citations and the geography of knowledge spillovers: a reassessment. Am Econ Rev 95(1):450–460
Wagner CS, Leydesdorff L (2005) Network structure, self-organization, and the growth of international collaboration in science. Res Policy 34(10):1608–1618
Wuchty S, Jones BF, Uzzi B (2007) The increasing dominance of teams in production of knowledge. Science 316(5827):1036–1039
Zander U, Kogut B (1995) Knowledge and the speed of the transfer and imitation of organizational capabilities: an empirical test. Organ Sci 6:76–91
Zucker L, Darby M, Armstrong J (1998) Geographically localised knowledge: spillovers or markets? Econ Inq 36:65–86

Not Invented Here: Technology Licensing, Knowledge Transfer and Innovation Based on Public Research

Guido Buenstorf and Matthias Geissler

Abstract Using a new dataset encompassing more than 2,200 inventions made by Max Planck Society researchers from 1980 to 2004, we explore the way in which inventor, technology, and licensee characteristics affect the commercialization of academic inventions. We find limited evidence suggesting that domestic and external licensees outperform foreign licensees and inventor spin-offs in the commercialization of academic inventions. Controlling for selection, spin-offs are indistinguishable from external licensees. Patented technologies and inventions by senior scientists are more likely to be licensed, but patent protection is related to lower commercialization odds and royalty payments.

1 Introduction

Throughout the developed world, public attention and policy initiatives increasingly focus on the transfer of knowledge from public research to the private sector. Following the Bayh-Dole Act of 1980 in the U.S. and subsequent legislative changes elsewhere, technology transfer has generally been accepted as a primary objective of universities and other public research organizations (cf. Mowery et al. 2001; Phan and Siegel 2006; Verspagen 2006). Notwithstanding the importance of alternative transfer channels (Cohen et al. 2002), commercialization of scientific results based on patents, licensing, and spin-off entrepreneurship has found particularly intensive policy attention as well as scholarly scrutiny (e.g.,

Reprinted from Journal of Evolutionary Economics 22(3), 481–511, Springer (2012).

G. Buenstorf (✉)
Institute of Economics and INCHER-Kassel (International Center for Higher Education Research), University of Kassel, Nora-Platiel-Strasse 5, 34109 Kassel, Germany
e-mail: buenstorf@uni-kassel.de

M. Geissler
Institute of Economics, University of Kassel, Kassel, Germany
e-mail: geissler@uni-kassel.de

G. Buenstorf et al. (eds.), *The Two Sides of Innovation*, Economic Complexity and Evolution, DOI 10.1007/978-3-319-01496-8_5,

Shane 2002; Lowe and Ziedonis 2006; Elfenbein 2007). Yet in spite of the increased emphasis on universities' intellectual property rights (IPRs) and IPR-based commercialization, little is known about the underlying processes of knowledge transfer.

Commercializing academic inventions is non-trivial because they are often far from being readily marketable. Prior work suggests that commercialization is complicated by uncertainty stemming from the early-stage character of most university inventions (Jensen and Thursby 2001), information asymmetry between inventor and potential licensee (Shane 2002), and also the non-codified nature of important elements of the knowledge base underlying the traded technology (Agrawal 2006). We lack conclusive evidence on how the challenges posed by these traits of academic inventions are related to inventor, technology, and licensee characteristics. For example, the relative commercialization performance of inventor spin-offs vis-à-vis external licensees is a contested issue (Shane 2002; Lowe and Ziedonis 2006). Other issues, including the effectiveness of international licensing, are largely unexplored.

Furthermore, prior empirical studies are based on U.S. data. In light of dissimilar academic traditions and substantial institutional differences, it cannot be taken for granted that their results generalize to other, in particular European, countries. We are not aware of any prior work studying the commercialization of academic inventions from Europe at the level of individual inventions. The dearth of empirical evidence is not surprising, given that, historically, European academic inventions were often owned by the inventors themselves (a practice known as the "professors' privilege"). European universities had little to license under these conditions.

In this paper, we begin to close this gap by studying the commercialization of inventions made by researchers at the Max Planck Society, Germany's largest non-university public research organization (PRO) dedicated to basic science. We exploit the fact that, in contrast to German university faculty, but similar to other German PROs such as the Fraunhofer Society, Max Planck scientists have never been covered by the "professors' privilege". Instead, the IPR regime that consistently has governed commercialization activities at the Max Planck Society since the 1970s closely resembles the one established in the U.S. by the Bayh-Dole Act, which has become the global template for dealing with academic inventions. The Max Planck Society, therefore, provides a rare opportunity to study the commercialization of European academic inventions in the now dominant institutional setting.

The dataset on the Max Planck Society's commercialization activities encompasses more than 2,200 inventions and about 700 license agreements providing for royalty payments over the time period 1980–2004. We use this dataset to analyze the way in which licensing and commercialization outcomes are affected by differences across inventors, technologies, and licensees that condition the relevance of information asymmetry and non-codified knowledge. Both the incidence and the level of royalties are utilized as measures of successful commercialization.

The present study aims to make the following specific contributions. First, we study licensing and commercialization outcomes across national boundaries. While less relevant in the U.S. context, licensing to foreign firms is a pertinent issue in the smaller and more open European economies, but has received little prior attention. Second, we also contribute new evidence to the unresolved issue of the effectiveness of inventor spin-offs as commercializers of academic inventions. Third, we analyze effects of technology and inventor characteristics on the outcomes of IPR-based technology transfer. In this context, we focus on the role of patent protection and inventor seniority. Fourth, the empirical analysis accounts for the possibility that non-random selection into licensing by different types of licensees may affect commercialization outcomes.

Our analysis indicates that information asymmetry and the difficulty of transferring non-codified knowledge are relevant in shaping the success of license-based technology transfer from public research, even though they cannot fully explain the empirical patterns. We find limited evidence suggesting that domestic and external licensees outperform foreign licensees and inventor spin-offs in the commercialization of academic inventions. However, these results are sensitive to varying model specifications. They moreover seem to reflect substantial effects of non-random selection into licensee types. Controlling for selection, spin-offs are indistinguishable from external licensees in their commercialization performance. Inventor seniority enhances the chances of technologies to be licensed, as does the presence and scope of patent protection. In contrast, patented inventions are less likely to yield successful commercial products.

The paper is structured as follows. The next section discusses the ways in which information asymmetry and knowledge transfer are relevant for commercialization activities in the empirical context of the Max Planck Society. A set of testable hypotheses is developed from the theoretical considerations. Section 3 introduces the methodology of the empirical analysis, while Section 4 describes data sources and the construction of our empirical measures. Results are presented in Section 5 and discussed in Section 6.

2 Technology Licensing at the Max Planck Society: Supply-Side and Demand-Side Considerations

2.1 The Role of Academic Inventions at the Max Planck Society

Scientists working in public research often make inventions that are suitable to provide the foundations of commercially viable innovations. However, developing products from these inventions and selling them in the marketplace is not part of the scientist's regular job. Instead, substantial "markets for technology" (Arora and Gambardella 2010) have developed for academic inventions. In these markets, licenses on academic inventions are sold to private-sector firms.

Since in most countries intellectual property rights in academic inventions are allocated to the inventors' employers (Lissoni et al. 2008), academic inventions are marketed by employers rather than inventors. To this purpose, most universities and PROs have set up dedicated entities known as technology licensing offices (TLOs). It is the representatives of these offices who actually operate in the markets for technologies. Inventors are nonetheless key players in the licensed-based commercialization of academic inventions. To assess their relevance in the present empirical context, a closer look at the organizational structure of the Max Planck Society is required.

The Max Planck Society is Germany's largest non-university PRO dedicated to basic research. At the end of the analyzed time period, it received almost 80 per cent of its budget from public, institutional funding (Max Planck Society 2008). The Max Planck Society's mission is to complement the German university system by taking up large-scale, interdisciplinary, or particularly innovative activities that are out of reach for individual universities or do not fit their organizational structure. To this purpose, the Max Planck Society operates about 80 individual institutes that are dispersed all over the country (plus three institutes located abroad) and cover a wide spectrum of research. Institutes are organized into three sections: the biomedical section; the chemistry, physics and technology section; and the humanities and social sciences section. Given its traditional orientation toward basic research, several patent-intensive fields of research, notably in engineering, are less important for the Max Planck Society than they are for universities and also for its more applied counterpart, the Fraunhofer Society.

In 2007, the Max Planck Institutes employed some 4,700 researchers (Max Planck Society 2008). While salaries are not much different from those paid at German universities, Max Planck researchers have no teaching obligations, and the availability of resources and equipment is generally better than in universities. In turn, Max Planck researchers are expected to attain academic excellence and to be international leaders in their fields. The performance of individual institutes is assessed by advisory bodies, and underachieving institutes can be restructured or even shut down. Individual-level performance is indicated by publications and their impact. Patent output is not generally used as a performance measure.[1] Likewise, given the relatively generous institutional funding enjoyed by the Max Planck Society, input-based performance measures (e.g., the amount of third-party funding attracted by researchers) are less relevant in the assessment of individual achievement than at other institutions. Again, in this regard, the Max Planck Society differs substantially from German universities and the Fraunhofer Society.

The internal organization of the Max Planck Society is unique. At the top of its scientific hierarchy are so-called Max Planck directors who enjoy particularly autonomous and powerful positions. New directors are recruited among the most successful researchers of domestic and foreign universities. The Max Planck Society currently has close to 300 active directors.

[1] Patents may have an indirect effect on the assessment of individual performance in fields where the scientific community values patents (Owen-Smith and Powell 2001). Prior research indicates that, in some cases, Max Planck researchers pursue patenting activities primarily to enhance their standing in the respective communities.

The Max Planck Society officially characterizes knowledge transfer through technology licensing as part of its objective to make research results socially relevant (Max Planck Society 2002). Just as the employees of private-sector firms, employees of the Max Planck Society are subject to the law on employee inventions ("*Arbeitnehmererfindungsgesetz*") requiring employees to disclose all inventions to their employer, and assigning the property rights in these inventions to the employer.[2] In case of successful commercialization of an invention, the inventor receives 30 per cent of all revenues from licenses and patent sales.

Max Planck Innovation GmbH, a legally independent subsidiary, is in charge of all activities related to academic inventions, patenting, and licensing. Max Planck Innovation was organized in 1970, originally under the name Garching Innovation. For the past three decades it has consistently focused on patenting and licensing activities.[3] Disclosure of inventions is actively solicited at the individual institutes. Patents are applied for if the invention is patentable and considered sufficiently promising, even if no licensee for the technology has been identified yet.[4] Technologies are marketed to both domestic and foreign firms. Systematic support and counseling of spin-off activities was taken up in the early 1990s, and spin-off numbers have strongly increased since then. At the end of the analyzed time period, overall licensing income contributed about one per cent to the Max Planck Society's annual budget (Max Planck Society 2008).

2.2 *The Supply of Academic Inventions: Incentives of Researchers and the Max Planck Society*

As noted above, the Max Planck Society's mission in the German academic system is to pursue excellence in basic research. This focus has repercussions on the incentives that Max Planck researchers have for making inventions: inventions are not directly relevant for the assessment of their research performance, and may even harm their career chances if they compromise the researchers' "traditional" output in terms of publications. Given these incentives, two characteristics help explain the large number of academic inventions at the Max Planck Society. First, scientists frequently make inventions as joint products of their research activities. (Think of instrumentation or lab equipment first used for the researcher's own use.) Second, in the use-inspired fields of basic research known as "Pasteur's Quadrant" (Stokes 1997), results can regularly be published in a scientific journal *and* be

[2] Before the "professors' privilege" was abolished in 2002, the IPR regime in place at the Max Planck Society differed from that of German universities. University researchers used to be exempt from the law on employee inventions. They retained the intellectual property in their inventions (cf. Von Proff et al. 2012).

[3] Following failed attempts at constructing and marketing prototypes, in-house commercialization of Max Planck inventions was given up in the 1970s and was never taken up again.

[4] In this regard, Max Planck Innovation's patenting policy thus appears to be closer to that of MIT than that of the UC system (cf. Shane 2002; Lowe and Ziedonis 2006).

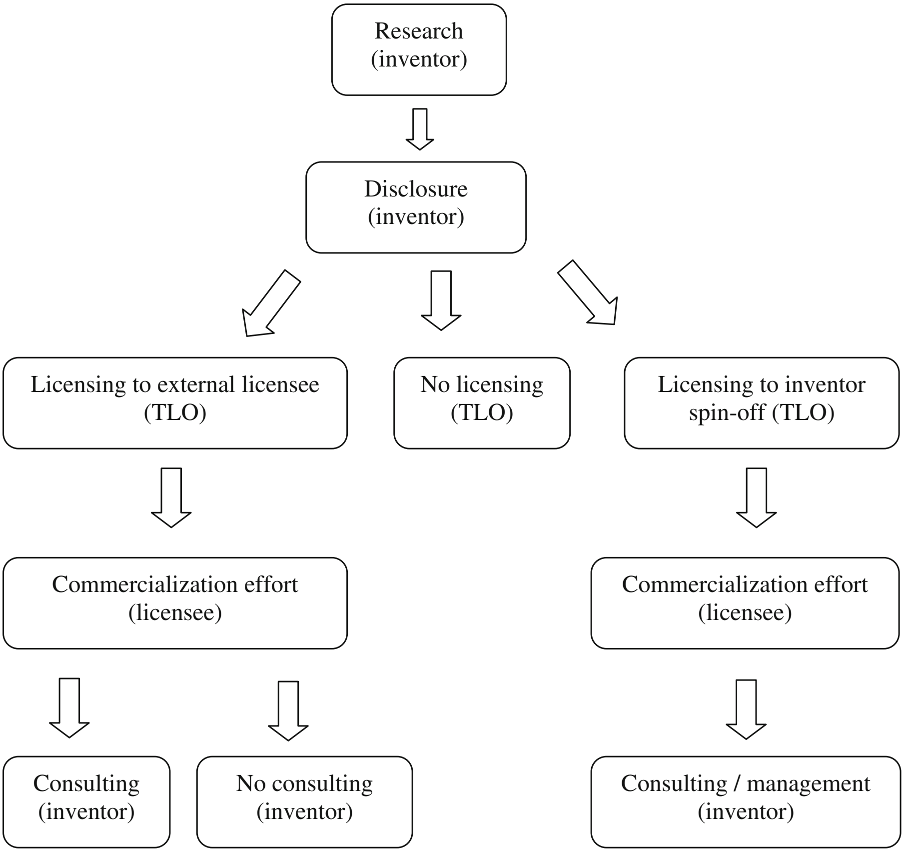


Fig. 1 Commercialization of academic inventions

applied commercially. For example, "patent-paper pairs" related to the same findings are widespread in the life sciences (Murray and Stern 2007).

The process leading from academic research to the successful commercialization of an invention is sketched in Fig. 1. Once an academic invention has been made by Max Planck researchers, it has to be disclosed to the Max Planck Society and becomes its intellectual property. As the Max Planck Society's agent, Max Planck Innovation then tries to license the invention, which at this time is frequently at an early stage of development. Prior research at U.S. universities has found that when inventions are marketed to potential licensees, the technology has often not advanced beyond the proof-of-concept or prototype stage (Jensen and Thursby 2001). Licensees accordingly need to engage in substantial further development efforts to obtain a marketable product.

Upon disclosure, an academic inventor has several options as to how to pursue the invention further. One possibility for the inventor is to focus on his (or her) academic research activities and to refrain from any further development of the invention. In this case, the fate of the invention entirely rests with Max Planck Innovation, which will try to find an external licensee for the (patented or unpatented) technology, often leveraging pre-established contacts to domestic or foreign firms. This may or may not be successful. For the Max Planck Society, successful licensing of technologies is attractive as a source of additional funding, and such licensing can be used to signal the societal relevance of its research activities to policy makers and the broader public. For Max Planck Innovation, successful licensing of inventions is attractive to signal its relevance to the Max Planck Society.

The inventor's second option is to become involved in the further development of the invention. If and when the invention is successfully licensed to an external licensee, he may choose to support the licensee's development efforts as a consultant. Such continued involvement of the academic inventor is often crucial for the successful commercialization of the invention, as it allows the licensee to draw on the inventor's non-codified knowledge (Agrawal 2006). On the downside, the inventor needs to allocate time to these consulting activities, which may have adverse effects on his research performance. In other words, continued involvement in the development of disclosed and licensed inventions comes with opportunity costs for the academic inventor. These opportunity costs will be the higher, the more complicated the interaction with the external licensee, and also the more valuable the research time of the inventor.

The third option available to the academic inventor is to pursue the commercialization of the technology himself by establishing a spin-off enterprise. Since the invention is owned by the Max Planck Society, the spin-off is required to license it back. The Max Planck Society supports spin-off activities of its researchers in various ways (Max Planck Society 2001) including cooperation agreements, access to the Society's infrastructure, and through assuming ownership positions in some of the new firms (substituting for upfront license payments). Ongoing involvement of researchers in the spin-off firms is allowed based on explicit consulting agreements. However, researchers may not assume management positions at the spin-off firm while maintaining their positions at the Max Planck Society (Max Planck Society 2002). An oft-observed pattern is that, within teams of academic inventors, the more senior partners (e.g., Max Planck directors) remain active researchers, while younger co-inventors (post-docs or doctoral students) join the spin-off management.

2.3 Adverse Selection and the Demand for Inventions from the Max Planck Society

The demand for inventions from the Max Planck Society ultimately depends on the ability of a potential licensee to make money from an obtained license. Whether or not a potential licensee expects a licensing agreement to be profitable hinges on a variety of factors. Three factors seem particularly pertinent: the ability to overcome information asymmetries in the negotiation of licensing agreements, successful knowledge transfer after a licensing deal has been made, and the command over sufficient capabilities and complementary assets actually to develop the technology further and to market it profitably. The first factor relates to the licensing stage of the technology transfer process and is discussed in the present subsection. The remaining two factors condition successful commercialization of licensed technologies and are the focus of the next subsection.

Technology markets tend to be thin; typically at best a few potential licensees exist for a particular technology, and licensing is based on small-numbers bargaining. Problems of asymmetric information are pervasive in these markets (Gallini and Wright 1990; Arora and Gambardella 2010) because, as opposed to technologies developed in-house, potential licensees of academic inventions lack in-depth knowledge of the underlying research. This limits their ability to assess the commercialization prospects of the invention, leading to problems of adverse selection.

Information asymmetry is minimized if a technology is licensed to a spin-off organized by its inventor(s). As regards external licensees, information asymmetries are expected to be more pronounced in licensing negotiations across national boundaries. Information is harder to obtain for foreign licensees, particularly if they do not come from countries speaking the same language. In addition, the design and enforcement of contracts is more difficult internationally.

The likelihood that a licensing agreement can be concluded with external licensees is enhanced by patents, which provide inventors with (imperfect) protection against being exploited by potential licensees. This enables Max Planck Innovation to disclose the invention more fully to potential licensees, thus mitigating the problem of adverse selection. Patents moreover signal that the invented technology conforms to an established standard of novelty, usefulness, and non-obviousness. The value of this signal is expected to be higher when information asymmetry is more pronounced. In addition, patents enhance the strategic value of a technology in blocking competitors' market access or in negotiating access to complementary technologies (Hall and Ziedonis 2001).

These arguments suggest that patents facilitate the licensing of academic inventions. They suggest that patents are more important in licensing technologies to foreign firms and external licensees. Based on similar considerations, Shane (2002) has suggested that spin-off licensing is a solution of last resort when attempts to find an external licensee have failed. Spin-off licensing would then be expected particularly when IPR protection is weak. However, patents are relevant as signals not only for potential licensees, but also for other transaction partners such as providers of external financing (Levin et al. 1987). This is particularly pertinent in the case of spin-offs, for which a substantial IPR base may be crucial to attract venture capital (Shane and Stuart 2002; Eckhardt et al. 2006). From this perspective, spin-off licensing may be even more dependent on the presence of patents related to an invention than licensing to external licensees. This is reflected in our predictions:

Hypothesis 1a: The likelihood that an invention is successfully licensed is enhanced by the presence and scope of patents related to the invention.

Hypothesis 1b: Patents have a stronger effect on the likelihood of licensing to a foreign firm than on the likelihood of licensing to a domestic firm.

Hypothesis 1c: Patents have a stronger effect on the likelihood of licensing to an inventor spin-off than on the likelihood of licensing to an external licensee.

Inventor reputation is another factor that helps overcome problems stemming from asymmetric information (Shane and Stuart 2002; Mora-Valentin et al. 2004). In this context, we expect that inventor seniority affects the likelihood of successful licensing negotiation as well as the probable type of licensee. Substantial prior empirical research finds positive correlations between inventive output and the quantity and quality of research output at the level of individual academic inventors (e.g., Azoulay et al. 2009; Breschi et al. 2008; Buenstorf 2009). Seniority, therefore, signals invention quality to potential licensees, which should increase their willingness to enter into a contractual agreement. In addition, inventor seniority enhances the visibility of academic inventions, which may further increase the likelihood of a successful licensing deal (Elfenbein 2007). Finally, if negotiations are mediated by a technology licensing office (as is the case in our empirical sample), it is likely that senior scientists have more influence on their employer institution than more junior ones. This is expected to increase further the chances that a licensing agreement is concluded.

As with patents, the value of the signal provided by seniority should be highest when information asymmetry is most pronounced, i.e. in the cases of foreign and external licensees. Again, we expect seniority to be relevant not only for finding external licensees, but even more so for securing finance (as well as other kinds of necessary resources such as first-round employees) in spin-off entrepreneurship. We accordingly conjecture:

Hypothesis 2a: Technologies (co-) invented by senior scientists are more likely to be licensed than those by more junior researchers.

Hypothesis 2b: (Co-) invention by senior scientists has a stronger effect on the likelihood of licensing to a foreign firm than on the likelihood of licensing to a domestic firm.

Hypothesis 2c: (Co-) invention by senior scientists has a stronger effect on the likelihood of licensing to an inventor spin-off than on the likelihood of licensing to an external licensee.

2.4 *Knowledge Transfer, Capabilities, and the Commercialization of Max Planck Inventions*

Adverse selection arises as a problem in negotiating licensing agreements because both parties have incentives to withhold information. In principle, asymmetric information may also give rise to moral hazard after a licensing agreement has been made, but this problem may be solved by contractual provisions in the licensing agreement (Jensen and Thursby 2001; Lowe 2006). If an agreement providing for sales-based royalties is entered into, inventors have an interest in successful commercialization. By contrast, even if both parties faithfully try to share their knowledge, substantial obstacles in communicating this knowledge typically have to be overcome after a licensing agreement has been reached. These obstacles derive from the nature of the relevant knowledge, which tends to be complex and imperfectly codified.

Agrawal (2006) argues that academic inventions often draw on multiple fields of knowledge. Potential licensees are unlikely to have prior knowledge in all these fields. Accordingly, their absorptive capacities (Cohen and Levinthal 1990) may be insufficient to understand fully information related to the invention, even if the inventor and/or the TLO disclose all their knowledge. In addition, relevant elements of that knowledge may be non-codified (even if they would in principle be codifiable), in which case they can be characterized as "latent" (Agrawal 2006). For example, knowledge that the inventor gained from failed and therefore unreported experiments is not normally accessible to an external licensee.

We expect that knowledge transfer between licensor (the academic inventor represented by his employer's TLO) and licensee is the more difficult the larger the "cognitive distance" (Nooteboom 1999) between both parties. "Cognitive distance" is not observable. However, in our sample, we expect cognitive distance to be larger for foreign licensees because language barriers and geographic distance complicate communication and post-agreement inventor involvement. Traveling is more costly in terms of time and money, and the transfer of non-codified knowledge (which presupposes frequent face-to-face interaction) is possibly less effective if national boundaries have to be crossed. We therefore predict the following:

Hypothesis 3: Inventions licensed to foreign firms are less likely to be commercialized successfully than inventions licensed to domestic firms.

As compared to external licensees, commercialization activities by inventor spin-offs are expected to benefit from facilitated transfer of non-codified knowledge, as spin-offs are more intimately familiar with the scientific background of the licensed invention. However, additional factors are likely to affect commercialization outcomes. Firms differ in their dynamic capabilities of integrating new technologies, which derive from the firms' prior activities and competences (Teece et al. 1997). In the present empirical context, we expect substantial differences in the kind and richness of capabilities possessed by external licensees, which are typically established firms active in a variety of markets related to the licensed technology, relative to inventor spin-offs that tend to be younger and smaller.

In addition, external licensees may be more likely to command substantial under-utilized complementary assets enabling them to benefit from innovation (Teece 1986). Shane (2002) stipulates that spin-offs are inferior in commercialization because they lack the required complementary assets. However, for their sample of licensed inventions from the University of California system, Lowe and Ziedonis (2006) find neither lower commercialization odds nor lower licensing income for spin-off licensees. This indicates that Shane's argument may be of secondary importance.

Finally, differences in the motivations of the different types of licensees may also affect observable commercialization outcomes. Given a smaller product portfolio, spin-off survival is typically more dependent on specific technologies than survival of established firms. Spin-offs consequently face stronger incentives for successful commercialization (Lowe and Ziedonis 2006). If spin-offs are cash-constrained, they should be less prone than established firms to license inventions for primarily strategic reasons, i.e. to block competitors from the access to the underlying technology or to enhance their negotiation position in contexts of "patent thickets" (Shapiro 2000; Hall and Ziedonis 2001). This would add to the likelihood that successful commercialization is observed in cases of spin-off licensing.

These considerations and the available prior evidence do not suggest a clear-cut ranking in the commercialization odds of external licensees and spin-offs. We therefore predict the following:

Hypothesis 4: Inventions licensed to inventor spin-offs do not differ significantly from inventions licensed to external licensees in their likelihood to be commercialized successfully.

Turning to technology characteristics, the relationship between patent protection and commercialization of academic inventions is likewise not immediately obvious. On the one hand, writing a patent application forces inventors to codify substantial parts of the knowledge underlying their inventions. This would be expected to help subsequent licensees turn the invention into a commercially successful product. At the same time, the above considerations regarding strategic licensing suggest that patented technologies may be less likely to be commercialized. Based on the assumption that the challenges of knowledge transfer are more relevant for the commercialization of academic inventions than purely strategic licensing, we predict the following:

Hypothesis 5: The presence and quality of patent protection related to an invention is positively related to its likelihood of commercialization.

Finally, successful commercialization of academic inventions may also depend on the seniority of the inventor(s). As with patents, two counteracting effects of inventor seniority on the commercialization odds of academic inventions seem plausible. On the one hand, as was argued above, prior research indicates that more successful researchers may also have inherently superior inventions. Inventor seniority would then be expected to be related to higher commercialization odds and higher royalty income. On the other hand, the more senior an inventor is, the higher are his opportunity costs of post-agreement involvement in the licensee's development efforts. *Ceteris paribus*, senior scientists are therefore expected to spend less time on developing their inventions, which will lower the likelihood of successful commercialization. In general, we expect the quality effect of seniority to outweigh the opportunity cost effect. This assumption informs our final hypothesis:

Hypothesis 6: Technologies (co-) invented by senior scientists are more likely to be commercialized than inventions made by more junior scientists.

3 Econometric Approach

We empirically analyze licensing and commercialization outcomes for the population of inventions disclosed by Max Planck researchers in the time period from 1980 to 2004. As detailed above, commercialization of academic inventions is a sequential process. In the first stage of licensing, we can construct outcome variables indicating the conclusion of a licensing agreement (or lack thereof) for each invention in the dataset. In the second stage, only the subset of licensed inventions is at risk of experiencing a successful commercialization outcome. Following the earlier work on U.S. academic inventions (Shane 2002; Agrawal 2006; Lowe and Ziedonis 2006), we measure successful commercialization by the incidence and level of sales-dependent royalty payments.

3.1 Likelihood of Licensing

Multinomial logit models are employed to analyze the likelihood that a given invention was licensed to a specific type of licensee. We estimate two sets of models, with the alternative outcomes being, respectively, licensing to a domestic versus foreign licensee, or licensing to an external licensee versus an inventor spin-off. No licensing is the reference outcome in both sets of models. A number of inventions were non-exclusively licensed or consecutively licensed to firms falling in both outcome categories (i.e, domestic *and* foreign; external *and* spin-off licensee). In these cases, we concentrate our attention on the first licensing agreement concluded for the respective invention. Right censoring issues are minimized by only analyzing inventions for which at least three years of licensing information is available and including measures of disclosure years in the analysis. (Academic inventions are mostly licensed in the first years after their invention.)

An endogeneity concern exists to the extent that potential licensees have been involved in the research leading to the invention. In our empirical context, this is the case for inventions based on collaborative research with industry partners constituting potential licensees. Problems of knowledge transfer should be less pronounced for these inventions. In addition, it could be true that an industry partner withdraws from cooperation even before an invention is arrived at because its assessment of the research is low. This should increase the average quality and thus the licensing odds of inventions based on collaborative research. However, as pointed out by Lowe (2002), the positive effect of collaborative research on the likelihood of licensing might be mitigated if, in the process of collaboration, industry partners acquired sufficient knowledge of the invention to render subsequent licensing unnecessary (or undesirable). Based on the available data, we cannot assess the importance of these concerns, but we can control for the fact that an invention is based on collaborative research in our empirical analysis.

3.2 Likelihood of Commercialization

Commercialization of licensed inventions is studied in three steps. First, we estimate a set of simple logit models with commercialization as the dependent variable, using the set of all licensing agreements as our sample, and estimating standard errors clustered by invention to account for multiple licensing of the same invention. As noted above, commercialization is defined as the existence of positive royalty payments. Obviously, this restricts the analysis to the subset of licensing agreements that provide for sales-dependent royalties. Second, we also analyze the commercial success of licensed inventions using the amount of royalties as the dependent variable. Royalties are censored at zero, which is taken into consideration by estimating Tobit models.

A shortcoming of both approaches is that they do not account for selection effects: Inventions licensed to different kinds of licensees may differ in their characteristics, and these differences may affect their subsequent commercialization odds. For example, it could be true that spin-off licensing is turned to when external licensees cannot be found, and that spin-offs therefore tend to license inferior inventions. Our analysis of licensing indicates that there are indeed substantial differences between the technologies licensed to different kinds of firms, which suggests that selection into the different kinds of licensing contracts (domestic versus foreign, spin-off versus external) is not random.

To test whether commercialization outcomes of different types of licensees are due to differences in observables affecting selection into licensee types, we interpret licensing to distinct types as treatments, and estimate how being treated affected the commercialization likelihood using propensity score matching. Specifically, two propensity score matching estimators are employed: in the first one, the treatment consists in being licensed to a foreign licensee. In the second one, licensing to a spin-off constitutes the treatment.

The intuition underlying propensity score matching is as follows (Rosenbloom and Rubin 1983, Heckman et al. 1998, cf. also Sianesi 2001, Wooldridge 2002, ch. 18). In non-experimental data, for each observation, only one outcome (here: commercialization success) is observed. If Y_{i0} denotes observation i's outcome without treatment, Y_{i1} denotes observation i's outcome with treatment, and $T \in \{0, 1\}$ denotes treatment, we would like to know the treatment effect $Y_{i1} - Y_{i0}$, but can only observe one of the two outcomes. If selection into treatment is nonrandom, the effect of treatment on the outcome cannot be separated from the selection effect in the data.

Propensity score matching uses the available information on individual observations to generate a counterfactual control group from the untreated observations, such that differences in observable characteristics are minimized between the treated observations and the members of the control group. The basic approach is to calculate the probability of receiving treatment for each observation based on its observable characteristics, using probit or logit models. This conditional probability is the propensity score, which is then used for matching the treated observations to similar non-treated ones. Under the (untestable) assumption that selection into treatment only depends on observables, the average effect of treatment can then be estimated at the population level. Specifically, both the *average treatment effect* (ATE), $E(Y_{i1} - Y_{i0})$, and the *average treatment effect on the treated* (ATT), $E(Y_{i1} - Y_{i0} \mid T = 1)$, can be estimated.

Various propensity score-based matching methods have been proposed. When large samples of non-treated observations are available, each treated observation can be matched to an "identical twin," i.e. a non-treated observation that is very similar in its propensity score, and the outcomes of both observations are then compared. Alternatively, each treated observation can be matched to a weighted average of untreated observations, where the weights are determined by how similar the propensity scores of the untreated observations are to that of the treated one. The latter approach is adopted below. We report results obtained by estimating propensity scores with logit models, using a Gaussian kernel for matching, where the weights of the untreated observations follow a normal distribution around the propensity score of the respective treated one. The estimations were performed using the *psmatch2* routine for Stata 9.0 (Leuven and Sianesi 2003).

4 Data

4.1 *Inventions*

This study is based on two sets of data made available by Max Planck Innovation. The first dataset contains all inventions disclosed by Max Planck researchers[5] from the early 1970s to 2004.[6] In total, it encompasses 3,012 inventions. Of these, 1,885

[5] Researchers employed on a scholarship basis, mostly Ph.D. students and international postdocs, are not subject to the German law on employee inventions (*Arbeitnehmererfindungsgesetz*). To the extent that these individuals made inventions without other Max Planck researchers being involved, they do not show up in the data.

[6] Our invention data end in February 2005 and include six inventions disclosed early in 2005. In the subsequent analysis, these are merged into the group of 2004 inventions.

Table 1 Inventions disclosed by Max Planck researchers between 1980 and 2004 and resulting licensing agreements

	Inventions		Licenses providing for royalties	
	All	Patented[a]	All	Patented[a]
Inventions	2,270	1,432 (1,387)		
Licensed (at least once)	744	536 (531)	717	503 (499)
Not licensed	1,526	896 (856)		
Commercialized			358	214 (211)
Licensed to domestic firms[b]	553	402 (398)	487	349 (346)
Licensed to foreign firms[b]	191	134 (133)	230	154 (153)
Licensed to external firms[b]	518	344 (342)	490	313 (311)
Licensed to spin-offs[b]	226	192 (189)	227	190 (188)

[a]Numbers for analyzed "patented-only" subsamples in parentheses
[b] In the invention dataset these refer to the type and region of the first licensee

resulted in at least one patent application.[7] The database includes the title of the invention, names and institute affiliations of its inventors, day of disclosure and (if eligible) patent application, as well as information regarding further use of the invention.

We restrict our empirical analysis to the 2,392 inventions disclosed in or after 1980. Earlier inventions are excluded for three reasons. First, the earliest entries in the inventions dataset are not consistently inventions by Max Planck researchers, since, at the time, Garching Innovation (the predecessor of Max Planck Innovation) was offering its services to a variety of other PROs and even some commercial firms, whose inventions then show up in our data. Second, the quality of the earliest data was below that related to later inventions. Third, Max Planck Innovation's commercialization strategy changed very little after a leadership change in 1979.

Another 122 inventions out of the 2,392 had to be dropped because essential data for our analysis was missing (24 cases) or the invention was not made at a Max Planck Institute (98 cases).[8] Accordingly, the final dataset used in the empirical analysis contains 2,270 inventions (cf. Table 1). Out of the 1,432 patented inventions in the final dataset, relevant data for the construction of variables in the analyses on "patented-only" subsample could not be obtained in some cases, which reduces the subsample to 1,387 inventions.

[7] In 141 cases, no patent information was found even though the inventions database identified them as patented. We suspect that most of these cases reflect cancelled applications. They are treated as not being patented in the subsequent analysis.

[8] This includes inventions coming out of temporary research groups and also, in a few cases, out of the Max Planck Society's central administration.

4.2 Licensing Agreements and Outcomes

We matched the inventions dataset with a second dataset assembled from Max Planck Innovation's licensing agreements. Our data on licensing agreements extend almost three years beyond the last disclosure date. In this way, right censoring issues for later inventions are minimized. In total 744 inventions in the dataset (536 patented inventions) have been subject to at least one licensing agreement. For each agreement, information is available about licensee name and address, the dates when the agreement was concluded and (possibly) terminated, contractual arrangements regarding fixed fees and royalties, as well as actual dates and amounts of payments as of 2007.

Two factors complicate the analysis of the licensing agreements: First, non-exclusive contracts may lead to multiple licensees for a single invention. Second, a number of licensing agreements cover multiple inventions. Because we are interested in the commercial potential of individual inventions (and use invention-specific control variables), we analyze all inventions covered by such "bundled" licenses separately and add an indicator variable denoting them in the empirical models. Payments from these agreements (if any) were split equally between the involved inventions. We thus deal with 1,014 invention-license pairs.

The presence of positive royalty payments is used as an indicator of successful commercialization. As noted above, commercialization outcomes can only be identified if licensing agreements provide for sales-dependent royalty payments. This is true for 717 invention-license pairs, 358 of which indeed yielded positive royalties (cf. Table 1). These numbers are comparable to U.S. institutions studied before. For example, Lowe and Ziedonis (2006) study 734 licensing deals closed by the UC system between 1981 and 1999, of which 188 led to positive royalty payments. We are also interested in the levels of returns from licenses. From the Max Planck Innovation files we identified annual royalty streams until 2007 for all contracts. As royalties are extremely skewed, we use the log of cumulative royalty payments in the empirical analysis.

4.3 Explanatory Variables and Controls

A central interest of the empirical analysis relates to the relative commercialization performance of different types of licensees. To study effects of international licensing, licensees were classified into domestic versus foreign according to the postal address given in the licensing agreements. Accordingly, German subsidiaries of foreign companies are classified as German licensees. This is in line with our primary interest in potential difficulties arising from information asymmetries and the transfer of non-codified knowledge, which we expect to depend more on the licensee's physical location than on whether or not the licensee is foreign-owned. Out of the 744 inventions that got licensed, 191 were licensed to a foreign firm.

Spin-offs among the licensees were identified on the basis of Max Planck Innovation's spin-off database. Here, 226 cases can be observed as opposed to 518 cases of licensing to an external licensee.

Information about patents related to the inventions in the dataset was obtained through a patent family search in *Depatisnet*, the publicly available patent search site of the German Patent Office (DPMA), using the patent applications listed in the Max Planck Innovation invention database as our point of departure. A simple indicator variable was first constructed that denotes inventions related to patent applications. In addition, within the subset of patented inventions we use the size of the patent family to account for differences in patent quality. Family size indicates the geographical scope of the IPR protection sought by the patent application and is widely accepted as a measure of patent quality (Harhoff et al. 2003). We employ a dummy variable indicating "triadic" patent families, including at least one application each at the European Patent Office and its Japanese and U.S. counterparts. We also experimented with the number of IPC classes and granted patents in the family as quality indicators, but they were less predictive. Finally, patent information is used to identify collaborative inventions. We define as collaborative all inventions that were not exclusively assigned to the Max Planck Society (i.e., those assigned either to the Max Planck Society and a private-sector firm or exclusively assigned to a private-sector firm).[9]

Senior scientist involvement in inventions is measured by the presence of one or (in rare cases) several Max Planck directors in the list of inventors. This is justified by the distinctive position directors have in the Max Planck hierarchy. We identified the directors using published sources (Henning and Ullmann 1998; Max Planck Society 2000) and information provided by the Max Planck Society's human resource department.

We control for discipline-specific factors with a dummy variable denoting inventions from the biomedical section, which accounts for 61% of all disclosed inventions. Controls are also included for inventions from the top five institutes in terms of the number of disclosed inventions. They include four institutes from the biomedical section and between them account for 42% of all disclosed inventions (45% of all licensed inventions). To capture time effects, we employ dummies denoting the year of invention disclosure.

Section and institute controls are the best indicator of the research field underlying an invention that we can develop for the full dataset including non-patented inventions. Table 2 provides more details on the composition of disclosure and licensing activities by sections and institutes. The table also distinguishes between patented and non-patented inventions. Inventions from the biomedical section

[9] Patent ownership is a restrictive measure of collaborative invention (Fontana and Geuna 2009), which is reflected by the comparatively small number of collaborative inventions we thus identified. We alternatively considered using information about collaboration from the Max Planck Innovation invention database. However, since the database is updated regularly and we do not have information about when the collaboration information was entered, we did not use it in the analysis based on endogeneity concerns.

Table 2 Descriptive statistics on license involvement by discipline, 1980–2004

	Not patented		Patented		
	Licensed	Not licensed	Licensed	Not licensed	Sum
Biomed section	148 (10.77%)	366 (26.64%)	403 (29.33%)	457 (33.26%)	1,374
- Top institute 1	36 (12.50%)	68 (23.61%)	85 (29.51%)	99 (34.38%)	288
- Top institute 2	20 (8.13%)	65 (26.42%)	80 (32.52%)	81 (32.93%)	246
- Top institute 4	14 (11.48%)	37 (30.33%)	35 (28.69%)	36 (29.51%)	122
- Top institute 5	10 (8.62%)	35 (30.17%)	37 (31.90%)	34 (29.31%)	116
- Other biomed section	68 (11.30%)	161 (26.74%)	166 (27.57%)	207 (34.39%)	602
Chem.-Phys.-Tech. Sec.	60 (6.70%)	264 (29.46%)	133 (14.84%)	439 (49.00%)	896
- Top institute 3	7 (3.70%)	61 (32.28%)	11 (5.82%)	110 (58.20%)	189
- Other CPT section	53 (7.50%)	203 (28.71%)	122 (17.26%)	329 (46.53%)	707
Overall	208 (9.16%)	630 (27.75%)	536 (23.61%)	896 (39.47%)	2,270

generally have a higher likelihood of being licensed. This holds both for patented and non-patented inventions, with the latter being less likely to be licensed throughout. Differences between the individual leading institutes within the biomedical section are less pronounced. The share of patented inventions is similar between both sections (62% in the medical section versus 64% in the chemistry, physics and technology section).

In the analysis of commercialization outcomes, we control for different experience of licensees by introducing an indicator for "serial" licensees showing up multiple times in our agreement dataset. Bundles of licensed inventions covered by individual contracts (cf. the above discussion) are also controlled for. Finally, four sectoral dummies denoting the broad area of activity of a licensee are included in the analyses of commercialization and royalty payments. These are based on NACE and SIC codes derived from the *LexisNexis* and *Hoppenstedt* firm databases as well as web searches. The dummies cover, respectively, manufacturing (SIC 20–39), wholesale/trade (SIC 50–51), services (SIC 70–89) and a catch-all variable including other industries, as well as licensees whose area of activity could not be reliably determined.

Descriptive statistics and correlations between the independent variables are shown in Tables 3, 4 and 5.

5 Results

5.1 Likelihood of Licensing

As detailed in the previous section, in terms of absolute numbers we observe more licensing to domestic firms (553 inventions) than licensing to foreign firms (191), and also more licensing to external licensees (518) than licensing to spin-offs (226). In this section, we employ two sets of multinomial logits to investigate the way in which the likelihood of being licensed by the different types of firms relates to

Table 3 Descriptive statistics

	All inventions (2,270 obs.)			Patented inventions (1,387 obs.)			Royalty contracts (all inventions, 717 obs.)			Royalty contracts (patented inventions, 499 obs.)		
	(mean)	(min)	(max)	(mean)	(min)	(max)	(mean)	(min)	(max)	(mean)	(min)	(max)
Director-inventor	.135	0	1	.180	0	1	.389	0	1	.437	0	1
Biomedical section	.605	0	1	.593	0	1	.778	0	1	.800	0	1
Patent application	.631	0	1	–	–	–	.702	0	1	–	–	–
Patent family size	–	–	–	5.416	1	120	9.062	1	74	9.062	1	74
Triadic family	–	–	–	.248	0	1	.383	0	1	.383	0	1
Collaborative invention	–	–	–	.208	0	1	.146	0	1	.146	0	1
Commercialization							.499	0	1	.423	0	1
Ln variable payments							4.761	0	19.109	4.152	0	19.109
Foreign licensee							.321	0	1	.307	0	1
Spin-off licensee							.317	0	1	.377	0	1
Bundle							.290	0	1	.363	0	1
Serial licensee							.760	0	1	.820	0	1

Table 4 Correlations between covariates I: likelihood of licensing (for patented inventions in parentheses)

2,270 obs. (1,387 obs.)	Biomed	Director-inventor	Patent	Patent family	Triadic family	Industry cooperation
Biomed	1.000 (1.000)					
Director-inventor	.168 (.201)	1.000 (1.000)				
Patent	− .013 (—)	.160 (—)	1.000 (—)			
Patent family	— (.148)	— (.218)	— (—)	— (1.000)		
Triadic family	— (−.036)	— (.138)	— (—)	— (.447)	— (1.000)	
Industry cooperation	— (−.121)	— (−.028)	— (—)	— (.112)	— (.194)	— (1.000)

observable characteristics of the respective inventions. Models 1–3 in Table 6 distinguish domestic from foreign licensees, whereas the corresponding Models 4–6 in Table 7 analyze licensing to inventor spin-offs versus external licensees. In both cases, we run models without (Models 1 and 4) and with (Models 2 and 5) controls for the top five institutes in terms of disclosed inventions, and also one model each (Models 3 and 6) with additional controls for invention quality. As they are based on patent information, these latter models are restricted to the subset of inventions related to patent applications.

We find that patented inventions are more likely to be licensed to either type of licensee. The estimated coefficient of the patent indicator is significant for domestic, respectively spin-off, licensees.[10] Among the patented inventions analyzed in Models 3 and 6, patent family size enhances the likelihood of licensing for all licensee types, consistent with Hypothesis 1a. Counter to Hypothesis 1b, the estimated effects of patent protection are statistically indistinguishable between domestic and foreign licensees. Consistent with Hypothesis 1c, the likelihood of licensing by spin-offs is more strongly related to the presence of patent applications than the likelihood of licensing to external licensees. This result supports the above conjecture that a strong IPR position is critical to inventor spin-offs dependent on external financing. Apparently, this role of patents is more important in our empirical context than either their capacity to help overcome information asymmetries or issues of strategic patenting, which would be expected to be more relevant for external licensees.

Hypothesis 2a predicted that inventor seniority is positively related to the likelihood of licensing across all licensee types. This finds support in the coefficients estimated for the indicator variable denoting (co-) inventions by Max Planck directors. As regards the nationality of licensees (Hypothesis 2b), differences in the effects of seniority are small and insignificant. In line with Hypothesis 2c, we find

[10] In an unreported simple logit model of licensing (irrespective of licensee type), the patent indicator is significant at the 1% level.

Table 5 Correlations between covariates II: likelihood of commercialization (patented inventions in parentheses)

717 obs. (499 obs.)	Foreign	Spin-off	Biomed	Director-inventor	Bundle	Exp. licensee	Patent	Patent family	Triadic family	Industry cooperation
Foreign	1.000 (1.000)									
Spin-off	− .245 (−.248)	1.000 (1.000)								
Biomed section	.180 (.105)	.111 (.162)	1.000 (1.000)							
Director-inventor	.132 (.063)	.213 (.249)	.192 (.148)	1.000 (1.000)						
Bundle	− .031 (−.050)	.265 (.231)	.016 (−.039)	.190 (.151)	1.000 (1.000)					
Serial licensee	− .153 (−.231)	.298 (.289)	.133 (.117)	.160 (.119)	.338 (.332)	1.000 (1.000)				
Patent	− .048 (−)	.202 (−)	.070 (−)	.152 (−)	.249 (−)	.212 (−)	1.000 (−)			
Patent family size	− (.184)	− (−.035)	− (.152)	− (.167)	− (.215)	− (.127)	− (−)	− (1.000)		
Triadic family	− (.102)	− (−.076)	− (−.038)	− (.005)	− (.203)	− (.112)	− (−)	− (.446)	− (1.000)	
Industry cooperation	− (−.115)	− (−.053)	− (−.090)	− (−.102)	− (.065)	− (.061)	− (−)	− (.027)	− (.129)	− (1.000)

Table 6 Likelihood of licensing I: domestic versus foreign (multinomial logit models)

	Model 1(all inventions)		Model 2 (all inventions)		Model 3 (patented inventions)	
	Domestic licensee	Foreign licensee	Domestic licensee	Foreign licensee	Domestic licensee	Foreign licensee
Patent application	.397*** (.121)	.247 (.184)	.400*** (.122)	.249 (.184)		
Patent family size					.101*** (.022)	.115*** (.023)
Triadic family					.271 (.210)	.389 (.293)
Director-inventor	2.966*** (.193)	3.051*** (.234)	3.133*** (.200)	3.127*** (.243)	2.888*** (.233)	2.776*** (.295)
Biomedical section	.535*** (.119)	1.485***[c] (.225)	.478*** (.145)	1.407***[c] (.251)	.605*** (.194)	1.068*** (.313)
Collaborative invention					.416**[c] (.184)	−1.110*** (.369)
Constant	−1.859*** (.227)	−4.087*** (.410)	−1.749*** (.233)	−3.829*** (.413)	−2.055 (.288)	−3.849*** (.495)
Time controls	Yes		Yes		Yes	
Top 5 institute controls	No		Yes		Yes	
Observations	2,270		2,270		1,387	
Log-likelihood ($p > chi^2$)	−1545.724 (.0000)		−1520.355 (.0000)		−931.239 (.0000)	
Pseudo-R^2	.169		.183		.238	

Standard errors in parentheses; *, ** and *** denote significance at the .10 .05 and .01 levels, respectively; [a], [b] and [c] denote differences between licensee types significant at the .10 .05 and .01 levels, respectively

that inventor seniority seems to play an even bigger role in the licensing to spin-offs than in the licensing to external licensees.

Two results related to the control variables are noteworthy. First, while inventions from the biomedical section are generally more likely to be licensed than inventions from the chemistry, physics and technology section, the difference is more pronounced for foreign licensees ($p < .01$ in Models 1 and 2). This pattern may reflect more developed markets for technology for biomedical inventions and/or the sectoral structure of the German economy, which is not specialized toward biomedical technologies. Second, inventions based on collaborative research are not generally more or less likely to be licensed,[11] but they are predominantly licensed to domestic firms and to external licensees. We interpret this as reflecting the composition of industry partners with which Max Planck researchers cooperate.

[11] In an unreported simple logit model of licensing (irrespective of licensee type), the indicator of collaborative inventions is insignificant.

Table 7 Likelihood of licensing II: external licensees versus spin-offs (multinomial logit models)

	Model 4 (all inventions)		Model 5 (all inventions)		Model 6 (patented inventions)	
	External licensees	Spin-offs	External licensees	Spin-offs	External licensees	Spin-offs
Patent application	.163 (.118)	1.075***[c] (.213)	.160 (.119)	1.072***[c] (.213)		
Patent family size					.100*** (.022)	.110*** (.023)
Triadic family					.447**[b] (.210)	−.122 (.290)
Director-inventor	2.761*** (.198)	3.427***[c] (.223)	2.936*** (.205)	3.535***[c] (.232)	2.643*** (.242)	3.264***[c] (.266)
Biomedical section	.606*** (.120)	1.179***[c] (.204)	.610*** (.144)	.973*** (.241)	.644*** (.197)	.966*** (.291)
Collaborative invention					.347*[b] (.191)	−.348 (.270)
Constant	−1.950*** (.237)	−3.803*** (.356)	−1.818*** (.242)	−3.648*** (.362)	−2.357 (.308)	−2.990 (.391)
Time controls	Yes		Yes		Yes	
Top 5 institute controls	No		Yes		Yes	
Observations	2,270		2,270		1,387	
Log-likelihood ($p > chi^2$)	−1530.722 (.0000)		−1505.714 (.0000)		−947.060 (.0000)	
Pseudo-R^2	.191		.205		.254	

Standard errors in parentheses; *,** and *** denote significance at the .10 .05 and .01 levels, respectively; [a], [b] and [c] denote differences between licensee types significant at the .10 .05 and .01 levels, respectively

5.2 Likelihood of Commercialization

To identify factors influencing the likelihood that licensed inventions are successfully commercialized, we begin by analyzing the odds of commercialization using logit models (Models 7–10 in Table 8), with successful commercialization measured by an indicator variable denoting licensing agreements that led to positive royalty payments. Second, the logged amount of royalties is adopted as an alternative measure of commercial success (Models 11–14 in Table 9). We finally assess potential effects of non-random selection into licensee types using propensity score matching (Models 15–18 in Tables 10 and 11).

Hypothesis 3 predicted that foreign licensees are less likely to commercialize a licensed technology. The evidence regarding this prediction is mixed. We consistently find lower commercialization likelihoods for foreign licensees, but three out of four estimated coefficients are only significant at the 10% level. In the models analyzing levels of royalty payments, the variable denoting foreign licensees is negative and sizable but never significant. The propensity score matching models indicate that the observed differences in the full sample of inventions (Model 15 in

Table 8 Likelihood of commercialization (logit models)

	Model 7 (all inventions)	Model 8 (all inventions)	Model 9 (patented inventions)	Model 10 (patented inventions)
Foreign licensee	−.365* (.197)	−.340* (.196)	−.608** (.288)	−.496* (.295)
Spin-off licensee	−.497** (.213)	−.542** (.218)	−.357 (.272)	−.471* (.280)
Director-inventor	.184 (.220)	.139 (.222)	−.004 (.290)	−.021 (.292)
Biomedical section	−.497* (.266)	−.517* (.266)	−.660* (.352)	−.746** (.356)
Patent application	−1.400*** (.266)	−1.389*** (.229)		
Patent family size			−.000 (.015)	−.001 (.015)
Triadic family			−.069 (.296)	−.165 (.299)
Collaborative invention			.469 (.325)	.583* (.332)
Bundled license	.595*** (.227)	.573** (.230)	.497* (.290)	.479* (.287)
Serial Licensee	−.014 (.236)	−.015 (.241)	−.000 (.339)	.031 (.359)
Constant	1.413*** (.428)	1.176** (.500)	.170 (.512)	−.539 (.683)
Time controls	Yes	Yes	Yes	Yes
Top 5 institute controls	Yes	Yes	Yes	Yes
Sector controls (licensee)	No	Yes	No	Yes
Observations	717	717	499	499
Log-likelihood ($p > chi^2$)	−407.548 (.0000)	−404.818 (.0000)	−265.922 (.0000)	−261.670 (.0000)
Pseudo-R^2	.180	.186	.218	.230

Clustered standard errors in parentheses; *, ** and *** denote significance at the .10 .05 and .01 levels, respectively

Table 10) are mostly due to selection.[12] In contrast, within the subsample of patented inventions Model 16 selection does not explain the lower commercialization likelihood of foreign licensees, as the average treatment effect on the treated is significantly negative.

Turning to the commercialization performance of inventor spin-offs, Models 7 and 8 find a significantly lower commercialization likelihood of inventions licensed to spin-offs. In Model 10, restricted to patented inventions, we obtain a marginally significant negative coefficient for the spin-off variable. Similarly, royalty payments realized by spin-off licensing are significantly smaller than those of external licensees in Models 11, 12, and 14. Propensity score matching (Models 17 and 18 in Table 11) indicates that these differences mostly reflect effects of selection into the alternative types of licensees. Without matching, the commercialization likelihood of technologies licensed to spin-offs is 10 to

[12] To obtain propensity scores, a logit model for the likelihood of being licensed to a foreign licensee was estimated first. We use a specification similar to Model 2. Kernel-based matching of treated and untreated observations was then performed (cf. also Section 3). The common support condition is satisfied for all reported propensity score matching models.

Table 9 Royalty payments from commercialization (Tobit models)

	Model 11 (all inventions)	Model 12 (all inventions)	Model 13 (patented inventions)	Model 14 (patented inventions)
Foreign licensee	−.811 (.817)	−.781 (.801)	−2.051 (1.255)	−1.639 (1.235)
Spin-off licensee	−1.929** (.858)	−2.163** (.862)	−1.727 (1.102)	−2.175** (1.102)
Director-inventor	−.105 (.886)	−.297 (.887)	−.512 (1.215)	−.540 (1.197)
Biomedical section	−1.880* (1.027)	−1.825* (1.008)	−2.648* (1.508)	−2.839* (1.489)
Patent application	−4.716*** (.794)	−4.576*** (.792)		
Patent family size			−.006 (.059)	−.002 (.059)
Triadic family			.407 (1.195)	−.147 (1.193)
Collaborative invention			1.988 (1.367)	2.501* (1.357)
Bundled license	1.787** (.880)	1.664* (.877)	1.307 (1.243)	1.119 (1.207)
Serial licensee	.481 (.842)	.485 (.860)	.273 (1.408)	.533 (1.443)
Constant	6.510*** (1.556)	5.547*** (1.792)	1.898 (2.280)	−.684 (2.842)
Time controls	Yes	Yes	Yes	Yes
Top 5 institute controls	Yes	Yes	Yes	Yes
Sector controls (licensee)	No	Yes	No	Yes
Observations	717	717	499	499
Log-likelihood ($p > chi^2$)	−1474.272 (.0000)	−1469.755 (.0000)	−907.864 (.0000)	−902.640 (.0000)
Pseudo-R^2 (ML)	.241	.250	.277	.291

Clustered standard errors in parentheses; *, ** and *** denote significance at the .10, .05 and .01 levels, respectively

14 percentage points lower than that of technologies with external licensees. When matched technologies are compared, this difference is reduced to 5 percentage points or less. None of the estimated treatment effects of spin-off licensing on commercialization is significant at the 5% level. Based on these findings, we fail to reject Hypothesis 4, which predicted indistinguishable commercialization outcomes for spin-offs and external licensees.

In both the logit and the Tobit models, the coefficient of the dummy variable denoting patented inventions is sizable and strongly negative, indicating that these inventions had lower commercialization chances than unpatented technologies. The proxies for patent quality are non-predictive throughout. Accordingly, both parts of Hypothesis 5—predicting both the presence and the scope of patents to be positively related to commercialization outcomes—do not find support in the empirical evidence. The same holds for the conjecture about inventor seniority Hypothesis 6, as the

Table 10 Likelihood of commercialization, domestic vs. foreign licensees (propensity score matching)

		Model 15 (all inventions, 717 obs.)			Model 16 (patented only, 499 obs.)		
		Unmatched	ATT	ATE	Unmatched	ATT	ATE
Treated		.457	.457		.359	.359	
Untreated		.520	.476		.451	.475	
Difference		– .063	– .020	– .075	– .091	– .115	– .081
S.E. (bootstrapped)			.042	.039		.051	.053
95% Confidence interval	Lower bound		– .085	– .131		– .212	– .181
	Upper bound		.069	.014		– .026	.037

Kernel matching (Gaussian kernel); standard errors obtained through bootstrapping (n = 100), bias corrected confidence interval for standard errors

Table 11 Likelihood of commercialization, external licensees vs. spin-offs (propensity score matching)

		Model 17 (all inventions, 717 obs.)			Model 18 (patented only, 499 obs.)		
		Unmatched	ATT	ATE	Unmatched	ATT	ATE
Treated		.405	.405		.362	.362	
Untreated		.543	.454		.460	.362	
Difference		– .137	– .049	– .050	– .098	– .000	– .020
S.E. (bootstrapped)			.045	.041		.049	.050
95% Confidence interval	Lower bound		– .135	– .128		– .093	– .116
	Upper bound		.016	.036		.096	.062

Kernel matching (Gaussian kernel); standard errors obtained through bootstrapping (n = 100), bias corrected confidence interval for standard errors variable denoting Max Planck directors among the inventors of a technology has no discernible impact on commercialization success.

5.3 Robustness Checks

A variety of robustness checks were performed to deal with limitations of the empirical analysis.[13] To check the robustness of the multinomial logit models, we also estimated the corresponding multinomial probit models. They yielded very similar results to those reported above. We also experimented with estimating hazard rate models that simultaneously analyze the incidence and timing of licensing events. We prefer the multinomial logits and concentrate on reporting their results below because time to licensing may vary across the technologies covered in our dataset, adding to the unobserved heterogeneity and possibly biasing the results of hazard rate models.

In the analysis of commercialization likelihoods, we estimated two-stage Heckman selection models (Heckman 1979) to deal with the issue that commercialization outcomes are only observable for the non-random sample of licensed inventions. To this purpose, we employed the indicator variable denoting (co-) inventions by senior researchers, along with the other variables employed in the licensing models reported above, to predict selection into licensing. Using inventor seniority as the instrument in the Heckman models is in line with the above empirical results showing that (co-) inventions by senior researchers are substantially more likely to be licensed, but not systematically related to commercialization. Second-stage results of the Heckman probit models are similar to the simple logit models reported above. The null hypothesis of independence between licensing and commercialization cannot be rejected in any of the models.

6 Discussion

In the present article, we studied technology transfer based on the commercialization of academic inventions from a major European non-university PRO, Germany's Max Planck Society. Due to peculiarities in the treatment of academic inventions in Germany before 2002, data on the incidence and outcomes of technology licensing from the Max Planck Society are available over an exceptionally long period of time. These data inform our econometric analysis, which is based on the full population of Max Planck inventions for the 1980–2004 period and takes into consideration that only the selected subset of licensed technologies is actually at risk of being commercialized.

We guided our empirical analysis by theoretical predictions based on notions of information asymmetry (for the first stage of licensing) respectively knowledge

[13] All unreported results are available from the authors upon request.

transfer from inventors to licensees (for the second stage of commercializing licensed technologies). Our empirical findings regarding these hypotheses were mixed.

We found that foreign licensees were less frequent than domestic ones. There is also substantial evidence suggesting they were less successful in commercializing inventions, even though the differences to domestic licensees were not consistently found to be statistically significant. To our knowledge, these are the first results for international licensing and commercialization of academic inventions.

We also found that, when controlling for non-random selection of inventions into licensee types, inventor spin-offs were indistinguishable from external licensees, both in their likelihood of commercializing academic inventions and in the level of royalties they generate from product sales. This result is consistent with the only directly comparable study by Lowe and Ziedonis (2006) for the University of California system, who did not find systematic differences between established firms and startups (in many cases inventor spin-offs). It indicates that spin-offs' advantages in access to inventors' non-codified knowledge, and possibly their stronger motivation to commercialize, enable them to overcome their disadvantages in terms of lacking complementary assets and organizational capabilities.

Academic entrepreneurship through inventor spin-offs is of substantial interest to policy makers. Our results suggest that spin-offs may indeed be a suitable channel for the commercialization of academic inventions and not just a second-best commercialization when external licensees cannot be found (as suggested even though not directly addressed by Shane (2002)). At the same time, our results do not support a privileged treatment of spin-offs relative to external licensees. We would rather interpret them as suggesting that an unbiased quest for the most suitable licensee in each individual case may be the best policy for universities and PROs operating under a Bayh-Dole-like IPR regime. This conclusion resonates with recent findings by Belenzon and Schankerman (2009), who showed that U.S. universities pursuing a selective licensing policy based on regional development objectives performed worse than those that did not.

A complex role of patents emerges from our results. Spin-off licensing is more strongly related to patents than licensing to external licensees, and inventions related to patents are less likely to be commercialized than those that are not. These findings point to an important role of patents as signals to external providers of capital and to the strategic licensing of patented inventions, which resonates with recent work indicating that the traditional interpretation of patents as devices allowing innovators to appropriate the returns of their R&D efforts is too narrow (cf., e.g., Shapiro 2000; Hall and Ziedonis 2001; Jaffe and Lerner 2004).

Inventors' academic status is in the focus of Elfenbein's (2007) study of licensing at Harvard University. Similar to our results, he found that inventions by more prestigious scientists were more likely to be licensed, while contractual provisions did not differ with inventor status. This is interpreted as indicating that inventor status operates mostly through increasing an invention's visibility to external licensees—an interpretation that seems difficult to reconcile with our finding that seniority has stronger effects on licensing by spin-offs rather than external licensees.

Further theoretical and empirical work is clearly required to better understand the process through which academic inventions are turned into commercial success stories. That our knowledge about the commercialization of academic inventions is so limited is partly due to the paucity of suitable data. However, this may be a temporary problem: given the Bayh-Dole-like policies adopted in many countries and the setup of TLOs at most universities and PROs, data availability is bound to improve over time. In this context, it is also noteworthy that the Max Planck Society, which has consistently been subject to a Bayh-Dole-like IPR regime since the 1970s, was among the pioneers of IPR-based technology transfer in Europe. As a consequence, while the generality of our results is limited because they refer to a single organization, their relevance is enhanced by the fact that the same kind of IPR regime now governs the vast majority of European universities and PROs.

Acknowledgements We are grateful to Max Planck Innovation GmbH for granting us access to their data. For helpful comments and discussions we thank Joerg Erselius, Astrid Giegold, Dietmar Harhoff, Bernhard Hertel, Evelyn Kaiser, Lionel Nesta, Alexander Schacht, Vangelis Souitaris and Dieter Treichel, as well as two anonymous reviewers. Wolfgang Ziegler and Sebastian Schmidt at the Patent Information Office of the University of Jena helped us assemble the patent database. Wolfhard Kaus provided valuable research assistance. This article is based on research done while both authors were at the Max Planck Institute of Economics. The usual caveat applies.

References

Agrawal A (2006) Engaging the inventor: exploring licensing strategies for university inventions and the role of latent knowledge. Strateg Manage J 27:63–79

Arora A, Gambardella A (2010) Ideas for rent: an overview of markets for technology. Ind Corp Change 19:775–803

Azoulay P, Ding W, Stuart T (2009) The impact of academic patenting on the rate, quality and direction of (Public) research output. J Ind Econ 57:637–676

Belenzon S, Schankerman M (2009) University knowledge transfer: private ownership, incentives, and local development objectives. J Law Econ 52:111–144

Breschi S, Lissoni F, Montobbio F (2008) University patenting and scientific productivity: a quantitative study of Italian academic inventors. European Management Review 5:91–109

Buenstorf G (2009) Is commercialization good or bad for science? Individual-level evidence from the Max Planck Society. Res Policy 38:281–292

Cohen WM, Levinthal DA (1990) Absorptive capacity: a new perspective of learning and innovation. Adm Sci Q 35:128–152

Cohen WM, Nelson RR, Walsh JP (2002) Links and impacts: the influence of public research on industrial R&D. Manag Sci 48:1–23

Eckhardt JT, Shane S, Delmar F (2006) Multistage selection and the financing of new ventures. Manag Sci 52:220–232

Elfenbein D (2007) Publications, patents, and the market for university inventions. J Econ Behav Organ 63:688–715

Fontana R, Geuna A (2009) The nature of collaborative patenting activities. Bocconi University, mimeo

Gallini NT, Wright BD (1990) Technology transfer under asymmetric information. RAND J Econ 21:147–160

Hall BH, Ziedonis RH (2001) The patent paradox revisited: an empirical study of patenting in the U.S. semiconductor industry, 1979–1995. RAND J Econ 32:101–128

Harhoff D, Scherer FM, Voper K (2003) Citations, family size, opposition and the value of patent rights. Res Policy 32:1343–1363

Heckman JJ (1979) Sample selection bias as a specification error. Econometrica 47:153–161

Heckman JJ, Ichimura H, Todd P (1998) Matching as an econometric evaluation estimator. Rev Econ Stud 65:261–294

Henning E, Ullmann D (1998) Wissenschaftliche Mitglieder der Max-Planck-Gesellschaft zur Förderung der Wissenschaften im Bild. Duncker & Humblot, Berlin

Jaffe AB, Lerner J (2004) Innovation and its discontents. Princeton University Press

Jensen R, Thursby M (2001) Proofs and prototypes for sale: the licensing of university inventions. Am Econ Rev 91:240–259

Leuven E, Sianesi B (2003) PSMATCH2: Stata module to perform full Mahalanobis and propensity score matching, common support graphing, and covariate imbalance testing. http://ideas.repec.org/c/boc/bocode/s432001.html

Levin RC, Klevorick AK, Nelson RR, Winter SG (1987) Appropriating the returns from industrial research and development. Brookings Pap Econ Act 1987(3):783–820

Lissoni F, Llerena P, McKelvey M, Sanditov B (2008) Academic patenting in Europe: new evidence from the KEINS database. Res Eval 17:87–102

Lowe RA (2002) Entrepreneurship and information asymmetry: theory and evidence from the University of California. Carnegie Mellon University, mimeo

Lowe RA (2006) Who develops a university invention? The impact of tacit knowledge and licensing policies. J Technol Transf 31:415–429

Lowe RA, Ziedonis AA (2006) Overoptimism and the performance of entrepreneurial firms. Manag Sci 52:173–186

Max Planck Society (2000) Handbook of scientific members. Munich

Max Planck Society (2001) Start-up companies: guidelines for scientists of the Max-Planck-Gesellschaft. Munich

Max Planck Society (2002) Guidelines for knowledge and technology transfer. Munich

Max Planck Society (2008) Annual Report 2007. Munich

Mora-Valentin EM, Montoro-Sanchez A, Guerras-Martin L (2004) Determining factors in the success of R&D cooperative agreements between firms and research organizations. Res Policy 33:17–40

Mowery DC, Nelson RR, Sampat BN, Ziedonis AA (2001) The growth of patenting and licensing by U.S. universities: an assessment of the effects of the Bayh-Dole act of 1980. Res Policy 30:99–119

Murray F, Stern S (2007) Do formal intellectual property rights hinder the free flow of scientific knowledge? An empirical test of the anti-commons hypothesis. J Econ Behav Organ 63: 648–687

Nooteboom B (1999) Innovation and inter-firm linkages: new implications for policy. Res Policy 28:793–805

Owen-Smith J, Powell WW (2001) To patent or not: faculty decisions and institutional success at technology transfer. J Technol Transf 26:99–114

Phan PH, Siegel DS (2006) The effectiveness of university technology transfer. Foundations and Trends in Entrepreneurship 2:77–144

Rosenbloom PR, Rubin DB (1983) The central role of the propensity score in observational studies for causal effects. Biometrika 70:41–55

Shane S (2002) Selling university technology: patterns from MIT. Manag Sci 48:122–137

Shane S, Stuart T (2002) Organizational endowments and the performance of university start-ups. Manag Sci 48:154–170

Shapiro C (2000) Navigating the patent thicket: cross licenses, patent pools, and standard setting. Innov Policy Econ 1:119–150
Sianesi B (2001) Implementing propensity score matching estimators with STATA. Prepared for the UK Stata Users Group, VII Meeting, London, May 2001
Stokes DE (1997) Pasteur's Quadrant. Basic science and technological innovation. Brookings Institution Press, Washington
Teece DJ (1986) Profiting from technological innovation: implications for integration, collaboration, licensing and public policy. Res Policy 15:285–305
Teece DJ, Pisano G, Shuen A (1997) Dynamic capabilities and strategic management. Strateg Manage J 18:509–533
Verspagen B (2006) University research, intellectual property rights and European Innovation Systems. J Econ Surv 20:607–632
Von Proff S, Buenstorf G, Hummel M (2012) University patenting in Germany before and after 2002: what role did the professors' privilege play? Industry and Innovation (in press)
Wooldridge JW (2002) Econometric analysis of cross section and panel data. MIT Press, Cambridge

Innovation in the Age of the Fuggers

Rolf Walter and Maximilian Kalus

Abstract The Fugger firm was one of the most diversified international, not to say global, players in the early modern era. The firms' success was based on a number of inventions, product novelties and social innovations, improvements in logistics. The Augsburg firm demonstrated an outstanding capability to combine different operating areas such as production, trade, logistics, services and banking. As a result, the Fuggers could secure competitive advantages in some new markets. A high degree of internationality, a dense network of agents and well-trained personnel, a most efficient logistics system and close personal contacts to the European courts and to curia and pope were some of the key elements of success of the Fugger company. Excellent human capital, which had often been educated in the core centers of the world economy such as Venice, Seville and Antwerp, was another core factor of the outstanding competitiveness of the firm. Large parts of the monetary capital of the Fuggers were transformed to cultural and social capital. This becomes evident by the existence of "social innovations" such as the Fuggerei in Augsburg.

1 Preliminary Remarks

To describe and analyze the Fuggers' innovations in the "Fugger city" of Augsburg has proved to be more complicated than expected. In general, the Fuggers were not so much inventors or innovators themselves, but rather adapted, imitated, distributed and combined novelty in an innovative way. As a consequence, the term "innovation" has to be considered in its widest meaning, including new markets, new products, and social innovation. Only by doing so can a number of innovations connected to the Fugger family be identified, which allows for an application of

R. Walter (✉) • M. Kalus
Friedrich-Schiller-University Jena, Jena, Germany
e-mail: r.walter@wiwi.uni-jena.de

G. Buenstorf et al. (eds.), *The Two Sides of Innovation*, Economic Complexity and Evolution, DOI 10.1007/978-3-319-01496-8_6,

Schumpeter's innovation theory. The same is true for the Welsers, Augsburg's second-largest company in the early sixteenth century.

Some remarks on the historical background may be instructive. Probably the most striking phenomenon in the early modern time is the material diversification and enrichment of economic and social life through contact with the New World, the "Las Indias". As Ernst Schulin put it once: It was a time in which humanity began to come to know herself. New cultures came up on the Europeans' horizon, especially regarding the Italian, Portuguese and Spanish seafarers. Weltanschauung and religion were fundamentally changed by the Reformation. The Renaissance emphasized rational thought and switched the focus from the next world to this one, from the collective to the individual. Art history provides us with some fine examples of this change at the turn of the sixteenth century, including Albrecht Dürer's self-portrayals or his individualized paintings of merchants such as Jakob Fugger.

Moreover, the names Fugger and Welser convey an acceleration and consolidation of networks during the transition from the Middle Age to the early modern time. Geographically speaking, both companies from Augsburg were among the furthest reaching ones worldwide. Thus, they were leading protagonists of the early globalization, the so-called proto-globalization.

While some travelled afar, seeking the alien by choice or passion, there were those who remained, whose existence and thoughts persevered, literally remaining bound to their native soil. In the early modern age, this still was the overwhelming majority of the population. Far more than 90 % of the people never moved farther than the imaginary radius of maybe 50 continental miles around their village, or had a clear notion of the world beyond. Although news networks slowly started to evolve during this time, the "world beyond" remained speculative. The opposite was the case with the deliberating and daring merchants of the Free Imperial City of Augsburg. Their far-reaching export of the well-known fustian was the corner stone of a transnational and transcontinental network, a *World Wide Web*—naturally reduced to the then discovered areas of the globe. During the age of the Fuggers, the known world was comprised of the Cape route discovered by Vasco da Gama, the islands of the Indian Ocean and the South China Sea, the West Indies, and coastal strips of a southern American continent. Also known, and of strategic prominence, were the islands of the Atlantic, which served as bridgeheads for the discoveries. The commerce with India was intense, while the commercial take-off of the American colonies was to begin in the second third of the sixteenth century. During this emergence of a world economy and the development of global networks, the great merchant-bankers of Augsburg, Jakob and Anton Fugger, as well as Bartholomäus Welser, played a leading role.

2 Business Based on Innovation

The ascendancy of the Imperial City of Augsburg and that of the Fugger Company in the late medieval period are closely related to an innovation in textile industry, namely the already mentioned fustian. Fustian is blended fabric, made of a flax warp and a cotton weft. For two centuries, fustian production was the driving force behind the Augsburg economy. The economic breakthrough can be placed around 1370, as many of the wealthy Augsburg families of the late Middle Ages and the early modern period were leading fustian distributors, producing the textiles in the putting-out system. Thus, the product innovation can be seen as having been connected to new ways of manufacturing, as well as to the expansion of procurement and sales markets. The Upper German fustian district proved to be one of the largest European industrial areas, stretching from Lake Constance to the Danube and Lech rivers. Merchants of the Swabian Imperial Cities of Nördlingen, Ulm, Ravensburg, Memmingen, Biberach, Konstanz and Kaufbeuren utilized their contacts with Venice to import cotton, and organized the regional textile production by selling the raw material, advancing credit and commercializing fustian as a commodity. Since 1395, fustian from Augsburg was sold on the Frankfurt fairs, and by the turn of the fifteenth century, further distributive channels through Cologne, Prague, Wroclaw, Cracow and Vienna were being opened up. It is possible to conceive these developments as a rapidly evolving industrialization of the Augsburg economy. Within this context, the Fuggers played an ever-growing role. But they were by no means the first ones to build up a commercial network and conduct textile, metal and financial business on a European scale. The Meuting company did the same 30 years before the Fuggers, not to mention the Große Ravensburger Gesellschaft (1380–1530), which sold linen, fustian and metal products throughout Europe (Italy, France, Spain, Cologne, Antwerp, Bruges, London, Vienna, Ofen, Wroclaw) in the late medieval era. They operated foreign branches called “Gelieger”, and imported cotton, as well as southern fruits and saffron (Kellenbenz 1977, 187 p).

When Hans Fugger moved from the village of Graben to Augsburg in 1367, the fustian boom had just begun and he could start his business thanks to some starting capital and two advantageous marriages.[1] He was able to participate in the commercial take-off of the late fourteenth and early fifteenth centuries. Since the 1440s, the Fuggers sold textiles at the important Pentecost fair of Nördlingen. On one occasion, they sold black fustian for 127 gold guilders. Their trademark, the trident, also emerged during that time. When the end of the fifteenth century brought temporary crises and a renewed boom, the Fuggers were prepared personnel-wise, financially and organizationally.

Several areas of commerce existed into which the Fuggers expanded and upon which they concentrated. Besides the production and the trade of quality textiles, the banking business proved to be the most important focus. At the end of the

[1] With Clara Widolf in his first marriage and with Elisabeth Gfattermann in his second one.

fifteenth century, the Fuggers' banking upgraded to the so-called high finance, advancing credit to princes, kings and the Habsburg emperors, Maximilian I, Charles V, and Philip II. In total, around a fifth (20 %) of the royal loans in the sixteenth century was alloted by the Fuggers (Carande 1943, 1949, 1967, passim). The Welsers, on the other hand, financed the French monarchy and maintained an important branch in the financial market of Lyon.

Another domain of the Fuggers was the Tyrolean mining industry, in which they possessed a strong position, although not a dominating one. Other companies were also present in Tyrol, such as Höchstetter, Baumgartner, Bimmel, Manlich, and Herwart. At the Falkenstein of Schwaz, a mine which produced silver, the Fuggers held around a 20 % interest. Consequently, they were not the only credit grantors of Archduke Ferdinand; the Haug-Langnauer-Linck firm, Joachim Jenisch and Wolfgang Paler also took part in this business.

Between 1565 and 1570, the smelting of the Tyrolean ore and the distribution of the resulting metal was conducted by the Jenbacher Company, a consortium of families just mentioned (Häberlein 2006, p. 109). Consortia and trusts were organizational forms used by the large companies in such cases.

Some markets experienced monopolistic tendencies, such as that of copper production. This was the third commercial domain of the Fuggers. Until 1546 they dominated the Slovakian mining industry (Kalus 1999, passim). *Besides bullion, copper was the pivotal mining product of this time, since its alloys bronze and brass were used in the production of canons and mortars, and thus was more important than iron* (Kalus 1999, p. 274).

In the expanding economy and monetary system of the late Middle Age, the demand for good copper and its alloys rose steeply, especially following the invention of black-powder guns and the mechanical wire traction. But "black copper" from the central European deposits was very brittle, because it contained a large portion of foreign matter, among it 0.6–1.8 % silver. This required the development of a metallurgic invention to divide both metals, the "Saigern". This process produced "Garkupfer" ("done copper") which was much more malleable than its black counterpart, and could be combined with calamine (zinc carbonate) to create brass, a raw material for wire-drawing mills. To divide silver from black copper, a large amount of lead and huge amounts of charcoal were needed. The process evolved between 1360 and 1450 (von Stromer 1995, p. 187).

The enormous output of the smelters of Neusohl (today Banská Bystrica in Slovakia) would not have been possible without the *saiger* process, because only through this could large mills be constructed and operated profitably. Neusohl with its large ore deposits was home of one of the main factories of the Fuggers and was led by up to three leading managers, called factors, at once.[2] The factory boasted a transport fleet of more than 160 coaches[3] that were used to transport coal, charcoal,

[2] After Anton Fugger decided to leave Neusohl in 1546, other entrepreneurs were quick to follow. These were the Manlich, Paler, and Weiß.

[3] The "Kotzje" had four spring-borne wheels. The coach was an invention of its time, hailing from the Hungarian town of Kocs.

and copper. Since the area was mountainous, the coaches had to be harnessed with up to eight horses. The Neusohl factory's stables contained 440 horses for the task. Moreover, in order to provision the mines and the mills, the Fuggers operated at least one dairy. Provisions were generally sold in the context of the "Pfennwerthandel" in which the company was the sole vendor of goods at the mines.

In Hungary, the Fuggers also traded oxen in large amounts.[4] This has to be perceived in connection with the mining business, since the mines needed enormous amounts of tallow to illuminate the tunnels. Moreover, the mining areas were important demand centers for leather, for both miners' clothing and buckets to drain the galleries (Westermann n.y., p. 273). The oxen's material was completely used up: Its horn was utilized to produce door handles, cutlery, etc., while the bones were bleached and turned to create devotional objects, combs, crosses and the like. Tallow was used to produce candles and soap, and to grease coaches. Last, but not least, oxen meat could be consumed. During the fifteenth and sixteenth centuries, hundreds of thousands of oxen were driven through central Europe. They tied up large quantities of capital, but needed no extra means of transportation. The mining districts of the Fuggers with their high demand for meat, leather and tallow, were natural consumer markets of oxen.

3 Improvements in Logistics

In order to orientate themselves, the merchants and captains of the sixteenth century used instruments made of wood or metal, such as the astrolabe, the quadrant, calipers, the compass, and the pocket sundial. Frequently, these were German-made. Other aids for the traveler included travelogues, books of navigation and descriptions of itineraries. Personal experience, confidence, imagination and an unerring instinct were absolutely vital for the traveler. The large *merchant bankers* needed to position themselves on a global scale, necessitating the perpetual accumulation of data, news and ideas from foreign lands. During the early modern era, the rationalization of time and space increased rapidly, as ever larger amounts of space were traversed in decreasing periods of time. The postal service[5] was economized through the use of relay systems. The Fuggers were likely to have been the first Germans to have used mounted relay messengers and thus helped to spread this innovation in the sector of logistics. The spread of news was accelerated

[4] Some evidence can be found in: Hermann Kellenbenz, Anton Fugger 1493–1560, Sonderdruck Weißenhorn 1993, p. 76; Allgemein: Ekkehard Westermann, Internationaler Ochsenhandel (1350–1750), (Beiträge zur Wirtschaftsgeschichte, vol. 9), Stuttgart 1979, passim; Wolfgang von Stromer, Wildwest in Europa—der transkontinentale Ochsenhandel der Frühen Neuzeit, in: Kultur & Technik, Zeitschrift des Deutschen Museums 3/1979, pp. 36–43.

[5] E.g. by the family of Tassis (Taxis) from Bergamo.

by establishing a larger number of relay stations for the same distance. Up to 260 km a day could be travelled by using hourly reliefs for the riders. Naturally, this was a decision as to how much capital and personnel to invest. By switching not only the riders on an hourly basis but also the mounts, around 360 km could be travelled in one day. The travel time between Augsburg and Venice, for example, could be cut down by one day in such a case. A merchant utilizing fast messengers conquered economic distance and made it hard for his business rivals to compete with him. Speed and accuracy of news were crucial for a profitable business. The Fuggers profited immensely from their relay system.

The fees of transcontinental news increased disproportionately with every hour of time saved en route. The "Geheimpapiere" (secret documents) of the Nuremberg/Augsburg based Paumgartner company around 1490 exemplify this: A message arriving from Venice to Augsburg in 4 days cost twice as much as one taking 5 days, and four times as much as one taking 6 days (Kellenbenz 1987, p. 287). Information speed was a function of the need of urgency, immediacy and importance of a piece of information. Advantages of a few days, or even hours, could determine the gain or loss of fortunes. Moreover, for far traders and large-scale entrepreneurs such as the Fuggers, the creation of a company-controlled relay system was a necessity, because they transferred not only commodities, but also money, bills of exchange and news. At the same time, an independent news service was more secure and dependable than hiring external messengers. Turnovers could be achieved by selling the service to third parties such as the emperor, the princes or the curia. Because of this, the Fuggers' news system became an important political factor.

In certain sections, the Fuggers utilized the imperial postal service (run by the Taxis), as well as those of the mercantile cities. As such, there was an interdependency between the different postal services. The Fuggers sold their service to the emperor, the pope and the Republic of Venice for good money, but also dispatched messages within other services. Augsburg and Nuremberg became Europe's leading communications centers. There even existed professional information mongers in these cities (Kempter 1936, n.p.; Kleinpaul 1921, n.p.; Sporhan-Krempel 1968, n.p.).

The turn from the middle to the early modern times brought along a media revolution. Besides the news systems, printed matter was the main carrier of this revolution. Around 1,500, half a century after the invention of Gutenberg, around 30,000 books had been printed. At the end of the sixteenth century, this number had increased to around 70–90 million (North 2000, p. 5). Printing made knowledge duplicable, portable and accessible at any time. As a consequence, the new sciences (alchemy, mathematics, botany, metallurgy, etc.) experienced an unsurpassed propagation, accompanied by a dissemination of literacy. The structural change of media brought along a three-tier communications system, comprised of book, pamphlet and newspaper. The book was mostly a means to secure and transport knowledge, whereas the other two contained opinions and news. In the last years of his life, Jakob Fugger frequently sent "neue Zeitungen" (recent news) to the duke of Saxony, for example (Häberlein 2006, p. 61). Therefore, not only innovative commodities appeared on the market, but the new media spread market developments, ideas and ideologies to the most remote areas. The nascent age of exploration and invention would not have been the same without the possibility to duplicate information in an easy way (Lehmann-Haupt 1966, n.p.). It should be noted that, among the early printers in Europe, a large number hailed from Upper Germany or the Rhineland (Otte et al. 1963/1964, pp. 129–162; Kellenbenz 1970a, b, p. 465 ff.).

4 New Markets

The international merchants both created and followed the paths and relationships of a nascent global economy. If one thinks of the shifts of "global economies" (as formulated by Fernand Braudel) being formed by an "invisible hand", it was the forward-looking and strategically minded merchant who had to adapt dynamically to this change. To someone named Bartholomäus, Jakob or Anton, the macroeconomic structure might have appeared to change only slowly, but he had to be intelligent and creative to achieve and to maintain a powerful market position and to calculate the risks of a market far from perfect information.

The international *merchant banker* had to be present at the centers of world trade. At the same time, he was pressed to find areas (both geographically and figuratively) in which he acted as *first mover*, being able to harvest rewarding rates of return and to create unique selling points. Such a change can be observed during the transition from the fifteenth to the sixteenth century in the geographic area of Italy and the Levante towards a region containing the Iberian coast of the Mediterranean, Seville, Lisbon up to Bruges and Antwerp. Often, the *global players* of their time followed such relocations by adapting their market orientation.

A perfect example for a shift from Venice[6] to Lisbon is the Welser company, and Lucas Rem. He was trained in the lagoon city between 1494 and 1495, learning languages used by merchants and mathematics ("rechnen in 5 ½ monet gar aus"). Moreover, he visited a school "da man biecher halten"—he was thouroughly trained in then (and still) up-to-date accounting techniques (Greiff 1861, p. 5). At that time, the company of the Welsers and Vöhlin imported indigo, cotton and all kinds of spices from Venice. This did not change when Bartholomäus Welser (V.) became head of the company in 1519, and the factory of Venice remained active throughout the sixteenth century.[7]

[6] The Welsers had been active in Venice since 1441 (Henry Simonsfeld: Der Fondaco dei Tedeschi in Venedig und die deutsch-venezianischen Handelsbeziehungen. Quellen und Forschungen, 2 Bde., Stuttgart 1887, here: vol. 1, pp. 232 ff.)

[7] Welsersches Familienarchiv, Signatur 305, zit. in: Großhaupt, Welser, p. 54. In 1588 Marx Manlich was the factor of the company in Venice.

4.1 The Welsers in Lisbon

As a result of the second crusade, Lisbon had been re-Christianized in 1147. During the turn of the fifteenth century, many internationally minded businessmen established themselves at the Tejo. The attraction of Lisbon increased after 1499, when war broke out between the Ottoman Empire and Venice, cutting Europe from Levantine trade for some time. For the Upper Germans, their center within the Republic of Marcus, the *Fondaco dei Tedeschi*, lost appeal. At the same time, Lisbon grew in importance and centrality.

Lisbon ascended shortly before 1500 when Vasco da Gama reached India via the seaborne Cape route. Da Gama had been back less than a year when Pedro Alvares Cabral was sent out on a second expedition to India. En route he discovered the coast of Brazil, which eventually led to a double expansion of the Portuguese crown. In its course, the relatively small kingdom of Portugal became a major world power which, in turn, roused the interest of international traders and financiers. The discoveries of Da Gama and Cabral raised hope in the large merchants and bankers of participating in the lucrative trade with the treasures of the Orient (spices) and that of Southern America (sugar and Brazilwood)—hope that was to be fulfilled soon. The Upper Germans were interested in the trade for several reasons. First, they were willing to pioneer the development of new markets with high yield expectations (and high risks). Second, they saw a chance to participate in a perfect barter business. The Welsers and other southern German merchants had access to large quantities of silver and copper, in addition to tin, lead and mercury. The *terms of trade* for such goods in Asia were remarkably good. Finally, the German investors were willing to take risks and could hope to receive protection from the Portuguese crown. The king was quite interested in the participation of large banking and merchant houses which could provide logistics, infrastructure and credit for the costly expeditions. The great merchants were keen to be granted exclusive privileges for the mercantile journeys to attain higher yields. This lowered the risks for the entrepreneurs, shippers and insurers who invested in the Portuguese expansion.

Among the *first movers* of the Portuguese Indian trade were the Welsers. Theirs was the first German company to invest in the Iberian Peninsula. In order to maximize economic possibilities, the firm followed a double strategy from the beginning: From Lisbon the East India trade was conducted, from Seville the path to the West Indies lay within reach (Walter 2006, p. 85 ff.; Pohle 2000, p. 99 ff.)

The base for the Welsers' mercantile presence was built by Lucas Rem. In September 1503, he bought "ain aigen herlich haus" (Greiff 1861, pp. 8 f.) ("an own splendid house") for his uncle Anton Welser and his associate, Konrad Vöhlin und Gesellschafter. Besides Rem, Simon Seitz and Scipio Löwenstein acted as representatives in Portugal. Lukas Rem stayed in Lisbon until 1508 as an employee of the Welsers and Vöhlin. His diary contains details on his commercial activities. He sold copper, lead, cinnabar, mercury in large quantities, and imported Flemish textiles and grain from the Netherlands, England, Brittany and the Baltic area. A large part of his trade was crown business, meaning that crown and court of Portugal were the main buyers. He merchandised spices in particular, but also oil, wine, figs, ivory ("helfentzän") and cotton. During the more than 5 years of his stay, Lucas and his brother Hans travelled to Maderia, the Azores, the Canary Islands (Tazacorte on La Palma 1509–1513) and northern Africa on behalf of the Welser-Vöhlin company. But his main objective was to convince the Portuguese to let Upper Germans participate in the expeditions directly. He achieved his goal on August 1st, 1504 when a contract was signed with the king to outfit three ships to be sent to India.

The drafting of the contract had been supported by an experienced German who was living in Lisbon, Valentim Fernandes, an important printer within the small kingdom. In the contract, the Welsers were allowed to trade with India directly, participating in expeditions sent from Lisbon. This privilege was extended to all German merchants who were willing to create a permanent factory in Lisbon and invest at least 10,000 ducats into the Portuguese trade. In return, they were granted toll reliefs on certain imported and exported products for 15 years, as well as the possibility to outfit their own ships to sail to India.

4.2 *The India Consortium of 1505/1506*

Certain foreign merchants were allowed to take part in the 1505/1506 expedition of Almeida. In order to finance the journey, a consortium of six Upper German and three Italian merchant houses was created. These were the families of the Welser, Fugger, Höchstetter, Imhoff, Gossembrot and Hirschvogel, as well as the Marchionni, Affaitadi (from Cremona) and Sernigi, andseveral smaller traders. The Welser-Vöhlin company participated with 20,000 *cruzados* (Portuguese Ducats), the Fuggers and Höchstetters with 4,000 each. The Imhoffs and Gossembrots invested 3,000 each, and the Hirschvogels spend 2,000 cruzados on the venture. The rest of 29,400 cruzados were raised by Italian companies (Ehrenberg 1898, pp. 195, 212; Kellenbenz 1970a, b, p. 319). Three of the 20 ships were outfitted by the German-Italian consortium which consisted of 14 "naos" and 6 caravels. The fleet set sail in March 1505 under the command of Francisco de Almeida. The consortium was to aliment and provision the crew members, who generally hailed from Portugal or her colonies. Only a few foreigners were allowed on board, one of them being the factor of the Welsers, Balthasar Sprenger from Vils am Lech.[8] (He sailed with the ship "Lionarda" and wrote down his impressions in a well-known essay called "Die Merfart" (The Sea Voyage), which contains detailed observations of the journey which ended on November 15th, 1506 (Pohle 2000, p. 205). Hans Mayr was another German

[8] Today Vils is located in Austria, very close to the border to the German town of Füssen.

onboard the "S. Rafael". Finally, on the "S. Jerónimo", the flagship of Almeida's fleet, there was the German Ulrich Imhoff, factor of the famous Nuremberg-based family Hirschvogel (Pohle 2000, p. 288).

The representatives of the consortium were allowed to buy as many goods as they were able to carry and pay for. On the return of the fleet, the merchandise had to be deposited in the Casa da India, the Portuguese institution which controlled the trade with Asia. After taxes had been paid, the goods could be sold as desired. The journey was a great success for the investors: The three ships of the consortium returned with a total of 12,000 quintals (hundredweights) of pepper. Even after the deduction of 30 % for the king, the pepper was sold for 168,000 cruzados. Compared to the total investment of 65,400 cruzados (Haebler 1903, p. 23), the profit amounted to around 150 %. Lukas Rem noted the immense return in his diary. The proceeds encouraged the German participants to invest in another expedition. Together with the Portuguese merchant Ruy Mendes, they outfitted three ships again and sent them to India in 1506. But the project failed. Only one of the vessels returned (Kellenbenz 1970a, b. pp. 319 f.). What the Germans had not considered was the high risk attached to the Asian trade during the early modern era. On the other hand, it also proved that the organizational pattern of the consortium shared and distributed risks among the investors. From a modern point of view, one would speak of the hedging function of a consortium as a means of managing risks.

5 Product Novelties and Innovation

The seaborne expeditions from the Iberian Peninsula led to new contacts in material culture. Dressing, eating and drinking habits were influenced by new goods coming from abroad. Everyday culture was enriched by noble metals, precious stones, pearls, and also by introducing new crop plants and beasts of burden. Know-how was transferred in both directions—to and from Europe. The Fuggers and Welsers took an important part of spreading such novelties. The Upper German entrepreneurs and mining business specialists were very interested in Southern America due to her rich deposits of gold and silver.

During Anton Fugger's lifetime, new processes in mining were invented and developed. One of those was the amalgamation to separate silver from other matter with the help of mercury. The Fuggers leased the so-called Maestrazgo rents to gain access to Europe's largest mercury deposit near the Spanish town of Almadén. The new technology was very labor- and capital-intensive, but promised an immense increase in the production of bullion. Technically the ore had to be crushed to fine sand. Mercury and potash was added, and the heap had to be shoveled periodically, adding water during the process. After 2–3 weeks, the amalgamation was finished and the silver had been absorbed by the mercury. Mercury and silver cold be separated easily by heating, although a substantial amount of mercury was lost in the process.

Around 1558, a new mercury deposit was discovered in Peru, and another in Huancavelica was located in 1564. At first, the findings did not solve the problem of mercury scarcity in the Americas, since the deposits were located at extreme heights. Only in 1571 did Pedro Fernández de Velasco perfect the chemical smelting process, a process that did not change for three decades afterwards.

For the next 10 years, there came around 7,000 quintal of mercury each year from Huancavelica. This enhanced silver production enormously. The wasteland of the Cerro Rico ("Rich Mountain") was settled and Potosí became the largest city of both Americas. In 1580, there lived around 120,000 inhabitants in the settlement, and by 1650, their numbers had risen to 160,000 (Bernecker et al. 1994, p. 428).

The mercury transport from Huancavelica to Potosí was difficult at best. The heavy liquid metal had to be transported in leather flasks by mules and llamas. The route was 1,200 km long and ran along mountain paths which reached up to 5,000 m above sea level. Although the mercury was very difficult to obtain, it was still a profitable business for the vice kingdom.

Besides their investment in bullion, the Fuggers were interested in exotic treasures, such as spices, medicine, perfumes, and the gemstones of the eastern Asian world. From Africa they imported gold, ivory and exotic animals, and from America pearls and other "Indian" exotica. Canada attracted them with furs, the Caribbean with salt, Brazil and Santo Domingo with exotic woods and sugar, Mexico with skins and indigo. From the tiny island of Cubagua near the Venezuelan coast, pearls were obtained. Completely new products were introduced, too: tobacco from Virginia, Venezuela and Brazil; cacao from Venezuela; and from the hinterland of the Guayaquil and Brazil, maize, potato, beans, tomatoes, peanuts and a variety of timber (e.g. balsa wood). These products were introduced to Europe through the expansion to the New World creating new markets, desires and consumer patterns, but also through economic restructuring (Kellenbenz 1971, p. 56 ff.). The indigo, for example, competed away the traditional European blue dye regions such as Erfurt (using woad).

The Welsers were a company willing to take great risks, and in 1528 they signed a contract with the Spanish crown to colonize Venezuela. They had built a factory in Santo Domingo, and wanted to use this platform to expand their markets to the rest of the Caribbean and to the American mainland and interior. Unfortunately for the Welsers, the venture turned out to be a grievous misinvestment. Anton Fugger, on the other hand, was more cautious. He had had a contract negotiated for Peru and Chile, apparently in order to acquire a better position to reach the "spice islands" in the Sunda archipelago, but the contract was never ratified.

Other business areas also profited from the expansion, such as the banking and financial sector. The bank of the Fuggers was active both in lending and in deposit-taking business. Deposits were accepted from private individuals and public institutions, and the bank paid interest or interest-like returns. Credits were given for trade, accommodation, loans, bonds, drafts and acceptances. In the name of the curia, they also conducted business for collection, e.g. with indulgences. Tax farming was another means of financing state businesses for collection, and the Fugger bank rented the taxes of the Spanish knight orders of Santiago, Calatrava and Alcántara, called Maestrazgos. For some time, the Fuggers also minted and issued coins on curial commission. In this business, they naturally profited from being miners of bullion and minters at the same time (Scheuermann 1929, n.p.; Walter 2003, n.p.).

6 Human Capital

The Fuggers were renowned not merely for their prowess as great merchants and bankers, but also for their highly skilled personnel. Business success was highly dependent on the selection and recruitment of well-qualified employees. Their experience, expertise, skill in negotiation and languages, as well as their deep knowledge of markets and fairs were crucial for efficient competition (Walter 2001, n.p.; Hildebrandt 1996, n.p.). A key position in a company was its "main factors", employees in central foreign markets. All the Fuggers— Jakob, Anton, Hans Jakob, Georg and Markus —C were keen to hire competent factors who were the most important protagonists within the complex network of the international financial and commodity markets. In Madrid during the latter half of the sixteenth century, this was Jobst Walther, followed by Christoph Hörmann in 1573, Thomas Miller and after, 1578, Hans Schedler. These individuals belonged to the elite of factors—renowned managers of their time. Management of human resources was critical for good business conduct, and so employees had to undergo rigorous screening and training (Bauer 1936, p. 34). Also business integrity was important in order to trust a factor.[9]

[9] Factors were senior employees, having been educated thoroughly. Apprenticeship often started at the age of 12, followed by training of 6 to 8 years. After this, the apprentice became a sales clerk who worked on a temporal employment for the company called "Verschreibung". The latter contained details on his rights and responsibilities, and the contract generally lasted for 6, 8, or—in rare cases—for 12 years (e.g. at the Welser company). At the end of the term, the employee received a "Verehrung", a bonus. They could then either continue to work for the company, switch employers or start their own business.

In general, factors were not allowed to lend money on their own account, or to receive gifts or other gratifications from third parties. Normally, they were had to work for their employer on an exclusive basis. Exceptions to the rules are known, such as the Neusohl-based factor of the Fuggers, Hans Dernschwam. By conducting business on his own account, he was able to accumulate quite a fortune. (Hermann Kellenbenz, (mit Götz Frhr. v. Pölnitz), Anton Fugger. Die letzten Jahre Anton Fuggers. Anton Fuggers Persönlichkeit und Werk, vol. 3. part II: 1550–1560 (Studien zur Fuggergeschichte 29), Tübingen 1986, p. 420).

The regulations under which factors had to work were very detailed and included rules on moral conduct. Debauchery were prohibited, e.g. playing cards, pomposity, "Kleidernarretei" (showing off clothes), illegitimate children, whoring, sodomy and the like. Hermann Kellenbenz, Die fremden Kaufleute auf der Iberischen Halbinsel vom 15. Jahrhundert bis zum Ende des 16. Jahrhunderts, in: L.c. (ed.), Fremde Kaufleute auf der Iberischen Halbinsel (Kölner Kolloquien zur internationalen Sozial- und Wirtschaftsgeschichte I), Cologne/Vienna 1970, pp. 265–376, The moral conduct of the employees became increasingly important. They were expected to lead a devout and respectable life, and had to be diligent, loyal and obedient in respect of the company owners. Moreover, they had to be able to keep secrets on the details of the business (Hildebrandt 1971, p. 166). As private persons, they were prohibited from gambling and bailing. Additionally, they were liable with all their fortune for damages to the company that were result of "unethical" conduct on their part.

The factors belonged to the elite of the international high finance, as they were "ambassadors" of Anton Fugger and his heirs. Frequently, they were connected to the Fugger family by kin- or friendship. In any case, they formed a precisely bordered social class. Naturally, the merchant princes were not interested in high fluctuation of their employees. It can be observed that the Fugger factors predominantly came from families that had been working for the company for generations, often in similar positions (Unger 1967, p. 225).

The factors were the leading managers of the companies. They carried businesses secrecies and professional, exclusive information on the company and its center (in Augsburg).[10] Each day they used this information to decide on tactical and strategic decisions.[11] They were close to the everyday market dealings, as they acted as the personal contacts of the company abroad. It was their task to deal with foreign market participants and business partners in person, and as such, they visited the fairs and other market events and often met high-ranking administrators or executive officers, such as notaries or royal officials.

Whenever large sums of money had to be transferred through fairs, factors and business agents were in charge. The financing of Charles V's election to Emperor in 1519 was conducted by factors, for example (Kellenbenz 1990, p. 65). It was natural for the factors and other leading managers of the companies to work together in such cases. In the election of 1519, factors of the Welsers and Fuggers cooperated closely.

Factors had to stay in continuous contact with royal courts. The Emperor steadily travelled throughout Europe with an entourage of far more than 1,000 people. This was costly, and he thus was in constant need of money. In autumn 1551, for instance, he wanted to leave Augsburg. The departure was to cost 76,000 ducats, which sum was raised in early October by the Fugger company ("Anton Fugger und Bruders Söhne", the so-called "Gemeine Handel"), the Welsers ("Hans Welser und

here: p. 320; Rudolf Ortner, Der Handlungsgehilfe, im besonderen der Faktor des süddeutschen Kaufmanns im 15. und 16. Jahrhundert, Munich 1932, p. 21.

[10] The necessity to keep the company's secrets was generally included in the work contracts and often continued after employment had been finished. Ortner, Handlungsgehilfe, p. 20; Ilse Lutzmann, Die Augsburger Handelsgesellschaft Hans und Marquard Rosenberger (1535–1560), Kallmünz 1957, p. 108.

[11] In this regard, they were similar to cameral servants of later times who also carried exclusive and secret knowledge of business affairs which they only shared as a last resort. More on this in: Kellenbenz, Hermann: Der Kammerdiener, ein Typus der höfischen Gesellschaft. Seine Rolle als Unternehmer, passim; Similarly on all factors: Rudolf Ortner, Handelsgehilfen und Reinhard Hildebrandt, Diener und Herren, passim.

Gebrüder") as well as Konrad Meuting. Gold and silver were used as collateral. The interest was fixed at 12 % and the amount was payable in autumn 1552 in the Spanish financial center of Medina del Campo (Kellenbenz 1990, p. 96).

The headquarters of the Fuggers in Augsburg and the international financial fairs experienced lively times when the Hapsburgs were about to give a celebration. Before Philipp married Mary of Tudor on July 25th 1554 in Winchester, the groom had let his secretary Erasso bargain with merchants of different nations to advance "mas o menos" (more or less) one million ducats for the festivity. The bulk of the expenses were paid by the Fugger factory of Antwerp lead by Matthäus Örtel (Kellenbenz 1990, p. 105).

Other large amounts of credits were lent to outfit and provision mercenary armies which were used to fight off the Turks or to campaign against the French king.

7 The Fuggers' Social Innovation

The Fuggers' innovations did not only concentrate on information networks and areas of commodity, production or logistics, but also included the social and charitable sphere. The Fuggerei is the best known example of such a social innovation. Another one was the so-called "Schneidhaus" (cutting house), a surgical institution which also belonged to the Fugger Foundations (Stein 2003, n.p.; Lieb 1958, p. 186[12]).

By founding the Fuggerei in 1521, Jakob Fugger became one of the larger caritative patrons of Augsburg, a housing estate within the walled Jakober Vorstadt ("suburb of St. Jacob") for pious Catholic, hardworking and blamelessly impoverished citizens of Augsburg. The deed decreed that every family had to pay a rent of one Rhenish Guilder. (Today the annual fee has dropped to 0.88 cent Euro, but in the sixteenth century the fee corresponded to the equivalent of a month's income of a day laborer) (Häberlein 2006, p. 148 ff.). Thus the rent was not just of a symbolic nature. Since the late Middle Ages, the city administrations began to distinguish between "decent" and "dishonest" paupers, provoked by a rising number of impoverished people within the cities. Only the "decent" paupers were deemed worthy of public support, which was generally carried out by religious or private foundations supported by wealthier citizens. The Fuggerei was not the only charitable institution within the city walls, but it was special through it sheer size—it boasted more than one hundred housing units, but lacked a communal center or public rooms, the reason being that this was the disciplinary character of the social sub-city which was surrounded by its own walls. The residents were supposed to enjoy neither gregarious meetings nor festivities nor idleness. They were rather supposed to lead a God-fearing and industrious life. The doors of the

[12] Thanks to Franz Karg of the Fugger-Archiv of Dillingen for the hint to that dissertation.

district were closed during the night, shutting out everybody who had left the taverns too late that evening.

For Jakob Fugger, the Fuggerei was one piece in a pious mosaic that also contained the chapel of St.-Anna (his tomb) and the foundation of a preacher's prebend in the church of St. Moritz in Augsburg. He regarded all three elements as a unit in order to save his soul. As successful merchant Jakob was in constant fear of his eternal soul, as he followed an often dubious occupation and did not have enough time to spend on a pious life. But the religious deeds were also a way to legitimize him within the city and raise the people's respect for his family. At the same time, his name became immortal as a generous patron of the Free Imperial City. Similarly, Jakob supported the Arts and music. Testimonials of this educating and learned patronage have survived in foundations of church organs, libraries and collections. These can be understood—to speak in the words of the French sociologist Pierre Bourdieu—as transformations of an economic towards a cultural capital (Bourdieu 1983, pp. 183–198).

8 Conclusion

During the fifteenth century, the Fuggers experienced a rapid economic ascension, and during the first two r thirds of the sixteenth century, they were able to expand their position by diversification and by scaling their output. The Fugger "conglomerate" commanded a capital of some millions guilders, and around 1560, it was likely to be one of the most potent companies on the globe. By 1545, the Fugger firm operated almost 50 branches within its European network that spanned from Seville to Riga and from Naples to London.

Jakob and Anton Fugger, as well as Bartholomäus Welser (V.), understood how innovatively to combine business fields and to generate and realize synergies. Their creative decision-takings are describable by the criteria of Schumpeter's theory of the dynamic entrepreneur. It was especially the capability to **combine** operating areas (production, trade, logistics, services, banking, etc.) in such a way that they enhanced each other and increased revenues where Fuggers and Welsers were brilliant. Jakob Fugger (the Rich) was able to gain access to and trust from Europe's most important political decision makers, the Habsburg kings and emperors, the Pope, the curia, the English king Henry VIII and countless administrators and dignitaries within the courts. Through this, the Augsburg-based family business could secure competitive advantages in quantitative (economies of scale) and qualitative (trust relationships, influence on jurisdiction, etc.) regards.

A high degree of internationality, a dense network of agents, informants and factors, a fast and reliable communications and logistics system, and close personal contacts to political and economic leaders were some of the key elements of success of the Fugger company. Their efficient commercial infrastructure was combined with excellent human capital, people who had often been educated in the Fondaco dei Tedeschi in Venice. Such constituents allowed the Augsburg merchant-bankers to develop new markets and—in some cases—even dominate them (e.g. the market of copper and lead).

The Fuggers also understood that one way to legitimize their fortune was by sponsoring social and cultural institutions. A large part of their monetary capital was thus transformed to cultural and social capital. This aspect becomes evident by the existence of "social innovations" such as the Fuggerei in Augsburg.

On the other hand, this close obligation towards the social elites, such as the commitment to the Spanish Habsburgs, contained dangers. The Habsburgs' continuous accumulation of debts carried immense risks for the main creditors, the Fuggers. Anton Fugger tried to diversify his risks. He gave credits to the English royal house which was more solvent than the Spaniards. He also attempted to have risks managed rigorously. The main factors tried to enforce the collection of debts by appealing to courts, for example. In the end, he could not escape the debt trap of the Habsburgs and had to write off several million guilders—the largest part of his fortune. His successors, Marcus Fugger and the "Georg Fuggerischen Erben" (Hildebrandt 1966, passim), were able to avert bankruptcy and stay relatively prosperous. The Welsers, on the other hand, went bankrupt in 1614, like many of the Augsburg companies after the state bankruptcies of Spain, Portugal and France, ending the "Golden Age" of the Free Imperial city.

References

Bauer C (1936) Unternehmung und Unternehmungsformen im Spätmittelalter und in der beginnenden Neuzeit. Jena, Gustav Fischer

Bernecker W, Buve R, Fisher J et al (ed.) (1994) Handbuch der Geschichte Lateinamerikas. Vol 1, Stuttgart

Bourdieu P (1983) Ökonomisches Kapital, kulturelles Kapital, soziales Kapital. In: Reinhard Kreckel (ed.), Soziale Ungleichheiten. Göttingen, Otto Schartz & Co, pp 183–198

Carande R (1943, 1949, 1967) Carlos V. y sus banqueros, vol 3, Madrid

Ehrenberg R (1898) Das Zeitalter der Fugger. Geldkapital und Creditverkehr im 16. Jahrhundert, vol 1, Die Geldmächte des 16. Jahrhunderts, Jena

Greiff B (1861) Tagebuch des Lucas Rem aus den Jahren 1494-1541. Ein Beitrag zur Handelsgeschichte der Stadt Augsburg. In: Jahresberichte des Historischen Kreisvereins für Schwaben und Neuburg, 26, Augsburg, pp 1–110

Häberlein M (2006) Die Fugger. Geschichte einer Augsburger Familie (1367-1650). Kohlhammer, Stuttgart

Haebler K (1903) Die überseeischen Unternehmungen der Welser und ihrer Gesellschafter. C.L. Hirschfeld, Leipzig

Hildebrandt R (1966) Die "Georg Fuggerischen Erben". Kaufmännische Tätigkeit und sozialer Status 1555-1620, Berlin

Hildebrandt R (1971) Diener und Herren. Eine historisch-statistische Untersuchung zur Entstehung der kaufmännischen Angestelltenschaft am Beispiel des Personals der großen oberdeutschen Handelshäuser 1450-1650 (Habilschrift, Ms. Vol. 1), Berlin

Hildebrandt R (1996) Diener und Herren. Zur Anatomie großer Unternehmen im Zeitalter der Fugger, in: Johannes Burkhardt (ed) Augsburger Handelshäuser im Wandel des historischen Urteils, Berlin, pp 149–174

Kalus P (1999) Die Fugger in der Slowakei (=Materialien zur Geschichte der Fugger, vol. 2, published by the Fugger-Archive), Augsburg

Kellenbenz H (1970) Die fremden Kaufleute auf der Iberischen Halbinsel vom 15. Jahrhundert bis zum Ende des 16. Jahrhunderts. In: L.c. (ed) Fremde Kaufleute auf der Iberischen Halbinsel, Cologne/Vienna, pp 265–376

Kellenbenz H (1970) Fremde Kaufleute auf der Iberischen Halbinsel (Kölner Kolloquien zur internationalen Wirtschaftsgeschichte I), Cologne/Vienna

Kellenbenz H (1971) Europa und die anderen Kontinente. Wirkungen der europäischen Expansion auf die Mutterländer. In: Spannungsfeld Ost-West. Strukturen und Persönlichkeiten, Munich

Kellenbenz H (1977) Deutsche Wirtschaftsgeschichte, vol. I. Von den Anfängen bis zum Ende des 18. Jahrhunderts, Munich

Kellenbenz H (1987) Die Entstehung des Postwesens in Mitteleuropa. In: Festschrift für Othmar Pickl zum 60. Geburtstag, Graz/Vienna, pp 285–291

Kellenbenz H (1990) Die Fugger in Spanien und Portugal bis 1560. Ein Großunternehmen des 16. Jahrhunderts, Teil I, Munich

Kempter K (1936) Die wirtschaftliche Berichterstattung in den sogenannten Fuggerzeitungen, in: Zeitung und Leben, 27

Kleinpaul J (1921) Die Fugger-Zeitungen 1568-1605 (Abhandlungen aus dem Institut für Zeitungskunde an der Universität Leipzig, vol. 4), Leipzig

Lehmann-Haupt H (1966) Gutenberg and the Master of the Playing Cards. Yale University Press, New Haven

Lieb N (1958) Die Fugger und die Kunst. Im Zeitalter der Hohen Renaissance (Die Krankenhausanstalten der Fuggerei) (Studien der Fuggergeschichte, vol. 14), Munich

North M (2000) Kommunikation, Handel, Geld und Banken in der Frühen Neuzeit (=Enzyklopädie deutscher Geschichte, vol. 59), Munich

Otte et al. (1963/1964) Die Begründer des deutschen Amerikahandels, in: Mitteilungen des Vereins für Geschichte der Stadt Nürnberg 52, pp 129–162

Pohle J (2000) Deutschland und die überseeische Expansion Portugals im 15. und 16. Jahrhundert (=Historia profana et ecclesiastica, vol. 2), Münster

Scheuermann L (1929) Die Fugger als Montanindustrielle in Tirol und Kärnten. Ein Beitrag zur Wirtschaftsgeschichte des 16. und 17. Jahrhunderts, Munich

Sporhan-Krempel S (1968) Nürnberg als Nachrichtenzentrum zwischen 1400 und 1700. Nürnberger Forschungen, Einzelarbeiten zur Nürnberger Geschichte, vol. 10, Nürnberg

Stein C (2003) Die Behandlung der Franzosenkrankheit in der Frühen Neuzeit am Beispiel Augsburgs, Stuttgart

Stromer W von (1995) Die Saigerhütten-Industrie des Spätmittelalters. Entwicklung der Kupfer-Silber-Scheidekünste zur "ars conflatoria separantie argentum a cupro cum plumbo". Technikgeschichte vol. 62 No. 3, pp 187–219

Unger E (1967) Die Fugger in Hall i. Tirol (Studien zur Fuggergeschichte 19), Tübingen

Walter R (2001) Human resources. Unternehmer und ihre Agenten auf den europäischen Märkten und Messen im 16. Jahrhundert. In: Istituto Internazionale di Storia Economica, Datini F (Ed) XXXII. Settimana di Studi, Fiere e mercati nella integrazione delle economie europee. Secoli XIII-XVIII, Prato/Firenze, pp 779–796

Walter R (2003) Das Silbergeschäft der Oberdeutschen in der Zeit Karls V. unter besonderer Berücksichtigung Lateinamerikas. In: Wolfgang Ingenhaeff/Johann Bair (ed.), Schwazer Silber – vergeudeter Reichtum? Verschwenderische Habsburger in Abhängigkeit vom oderdeutschen Kapital an der Zeitenwende vom Mittelalter zur Neuzeit, Innsbruck, pp 241–256

Walter R (2006) Geschichte der Weltwirtschaft. Eine Einführung, Cologne

Westermann E (n.y.) Forschungsaufgaben des internationalen Ochsenhandels aus mitteleuropäischer Sicht. In: L.c. (ed) Ochsenhandel, pp 261–288

Part II
Innovation: Induced Structural Change, Coping and Normative Assessment

Agent-Based Modelling of Novelty Creating Behavior and Sectoral Growth Effects—Linking the Creative and the Destructive Side of Innovation

Frank Beckenbach, Maria Daskalakis, and David Hofmann

Abstract For grasping the relationship between novelty creating activities of agents and growth of economic aggregates, a multi-level approach is suggested. The first level specifies the triggering conditions for novelty creating activities for the agents, i.e. firms. Here the behavioral elements and the modes of actions for the firms are portrayed using an agent-based approach (Section 2). On the second level, the consequences of successful innovations and imitations in a given sector of economic activities are dealt with (Section 3). This depends on the frequency of successful novelties and on the way they diffuse in that sector. We use an agent-related functional approach, applying difference equations for depicting the stylized facts of the diffusion dynamics. Only if these different levels of economic dynamics are distinguished as well as related to each other, is it possible to derive aggregate effects of novelties for the whole economy. This will be done by way of computer simulations (Section 4). Conclusions are drawn in Section 5.

1 Introduction

Obviously novelty creation[1] is a two sided process: it is 'creative' and 'destructive' at the same time. The background for this two-sided nature is that novelty creation has to overcome obstacles and, if it succeds in doing so these obstacles will be

Reprinted from Journal of Evolutionary Economics 22(3), 513–542, Springer (2012).
This research was supported by the Volkswagenstiftung. Comments by an anonymous referee and programming assistance by Ramón Briegel are gratefully acknowledged.

[1] In this article, we distinguish the notions of "novelty creation", "imitation" and "innovation". By the generic term "novelty creation" we integrate all aspects of the creation of new products in firms. This encompasses *radical product innovation* (simply called "innovation" in the sequel) as

F. Beckenbach (✉) • M. Daskalakis • D. Hofmann
Section Environmental and Behavioural Economics, University of Kassel, Nora-Platiel Str. 4, 34109 Kassel, Germany
e-mail: beckenbach@wirtschaft.uni-kassel.de

G. Buenstorf et al. (eds.), *The Two Sides of Innovation*, Economic Complexity and Evolution, DOI 10.1007/978-3-319-01496-8_7,

eroding therefrom, generating some individual or social costs. The main focus of this elaboration is on the obstacles for novelty creation, on the individual level in terms of routines, and on the sectoral level in terms of sticky demand both indicating well established and learned forms of behavior. Hence, the main question to answer is: What are the driving forces in favor of novelty creation and under what conditions are these driving forces strong enough to overcome the inertia of established behavior?

To assume that there is a given frequency of novelty creation and simply to suppose diffusion effects in sectoral terms would only be a partial answer to the question. Due to the inherently uncertain nature of the outcome of novelty creation, the conditions favorable for its occurrence have to be included in the analysis. Furthermore, the type of initial novelty creation (e.g. its individual or co-operative nature, or the distinction between innovation proper and imitation) is important for the dynamics of its social dissemination. Hence, a fully fledged concept of novelty-driven aggregate growth has to include

- the triggering conditions for novelty creation,
- the type of novelty creation and
- the dynamics of sectoral dissemination (diffusion).

Obviously, such an analysis is at least about two levels of economic activities and their interrelation.

This contribution tackles such a multi-level analysis. As regards the triggering conditions for novelty creation, a special emphasis is put on behavioral factors. Whereas external triggers in terms of scientific insights, environmental conditions or political regulations are thoroughly investigated (cf. Clarke and Weyant 2002 for a survey), the internal behavioral trigger elements are under-represented in the literature on novelty creation in general (exceptions: Shackle 1961; Scitovsky 1976; Witt 1987; March 1994; Morrison and Potts 2009) and especially in the literature on growth.[2] This is the more surprising given that the more attempts have been made to endogenize the explanation of growth by integrating the dynamics of novelty creation into the realm of growth theory. Specifying the behavioral trigger conditions for novelty creation (in terms of personal traits, on one side, and of individual variables on the other side) is the main task on the micro-level for analyzing the emergence of growth.

Against this background, this contribution aims to *first*, shed light on the micro-foundation for novelty creating activities, and *second*, show the interplay between the micro-level of firms and the sectoral level of economic activities. On both levels, novelty creation is not simply conceptualized as an add-on to given activities

well as, according to the definition of OECD, Eurostat (2005, p. 46), "imitations" which comprise innovations that are new at the level of the firm, but not for the market. Our definition of novelty creation takes a broader scope than the definition provided by Witt (2009, p. 312), who refers to fundamental novelty. On the other hand, our notion is narrower than Witt's as we refer only to product innovations. (These may involve physical products and/or services.)

[2] The so-called endogenous growth theory is the most prominent example for such an endeavor.

(which would imply a too simplistic view of the growth effect), but rather as an overcoming of given routinized activities and/or established products. Hence, specifying the devaluation or even destruction (in economic terms) as the other side of novelty creation is the *third* aim of this contribution. *Finally*, it will be shown how at least some basic stylized observations can be confirmed by such an analysis of the growth process: novelty creation as a temporary activity of a changing part of the whole economic population and persistent heterogeneity of actors in terms of behavior and economic performance.

In pursuing these aims, we generally follow the path of evolutionary thinking (Silverberg and Verspagen 1995, p. 4; Dawid 2006, p. 1243), taking into account that a twofold property of economic activities is at the core of such an approach: on one side, they are characterized by well established procedures and routines being backed by past experiences, while, on the other side, these activities are taking place in a dynamic selection environment urging the actors from time to time to search for new knowledge and corresponding new ways to act. Hence, on the individual as well as on the sectoral level, a breaking of routines and—as a possible consequence—a change of product properties as well as of exchange relations is occurring. In that sense, the creation and dissemination of knowledge is an important driver for economic evolution. According to this dynamic property, the actors in their entirety consist of different components as regards their behavioral status. They form a population the composition of which is changing over time. The most important hypothesis of an evolutionary approach to growth is that this time-dependent composition of the population as regards novelty creation mainly explains the time path of observable aggregates.

The sectoral growth model we are suggesting can be classified as a model in the tradition of Nelson and Winter (Nelson and Winter 1982; Chiaromonte and Dosi 1993; Silverberg and Verspagen 1994). In these models, the expenditure for novelty creating activities of firms is assumed to be the most important control variable. The experienced performance in the market competition is seen as the main factor influencing the setting of this control variable. The corresponding search activity of the firms takes place in a 'bounded' way (normally specified as 'local' search) resulting from the limited knowledge about the search space in which the firms are operating. Firms that are successful in such a search process have a comparative advantage over less successful firms. They gain additional profits and investment opportunities, both increasing their survival probability in the market.

But there are also a lot of features distinguishing the suggested model from the standard Nelson/Winter approach. It has a strong behavioral foundation in that the ability of firms to act is conceptualized in different modes of action.[3] According to the above mentioned twofold property of economic activities, every novelty creation has to overcome the routine mode to act and takes place in a specified manner

[3] A mode of action is defined by a particular way to care about information, to discriminate between alternatives and to link different practical operations.

as individual innovation or imitation or as a cooperative innovation.[4] These different modes to act are fed by different behavioral forces, being composed of (constant) personal traits and (variable) factors on the individual level. In that sense the occurrence of the intention to create a novelty as well as to pursue it in a specific mode of action is endogenized in behavioral terms in the suggested model architecture. This accomplishes the attempts in the literature to model novelty creating agents being endowed with different modes to act, in that the selection of these modes of action is now not merely determined by a given probability distribution (cf. Fagiolo and Dosi 2003) or derived from sectoral performance requirements (cf. Dawid et al. 2001; Llerena and Oltra 2002). Furthermore, in the model suggested here, the different modes to create novelty encompass different forms of knowledge acquisition and different effects of these knowledge augmenting procedures on sectoral growth. Due to this focus, the technical dimension of evolutionary change is not explicitly dealt with.[5] Rather, the latter is conceptualized in a disembodied manner (Silverberg and Verspagen 1995, pp. 4) in assuming that additional knowledge—as an ability to improve product quality—is tantamount to an access to additional market potential. This market potential is determined on the individual agent-related level. According to the usual spreading dynamics of information in the diffusion process (cf. Rogers 1995) it is assumed that this market potential for an improved product unfolds in an s-shaped manner in the course of time. Furthermore, the access to this market potential is influenced by agents imitating this improvement and by agents innovating later on, inducing a partial substitution of the demand for older products. Hence, the firm's overall access to the market potential is determined by different time-dependent layers of knowledge. Therefore, the model implicitly incorporates vintage effects and product differentiation.[6]

To summarize: for grasping the relationship between novelty creating activities of agents and growth of economic aggregates, a *multi-level approach* is suggested. The *first* level specifies the triggering conditions for novelty creating activities for the agents (i.e. firms). Here, the behavioral elements and the modes of action for the firms are portrayed by using an agent-based approach (Section 2). On the *second* level, the consequences of successful innovations and imitations in terms of their dissemination effects in a given sector of economic activities are dealt with

[4] In the given context, 'individual' means 'one firm' and 'cooperative' means at least two firms interacting directly.

[5] Apart from that tackling a technologically embodied change is not an easy analytical task for which a well defined solution recipe exists (Silverberg and Verspagen 1995, p. 9).

[6] Product innovation was introduced in a Nelson/Winter framework for the first time by Gerybadze (1982). Recently, Dawid and Reimann 2010 used an approach similar to the one suggested in the present contribution in a model on product diversification. In their model, the demand potential of a sub market is functionally dependent on the firms innovative efforts (measured in terms of r&d investments). This potential determines the attractiveness of a product variant. The attractiveness of each product variant changes according to a function in which time enters as an independent variable.

(Section 3). These depend on the frequency of successful novelties and on the way they diffuse in that sector. As regards the supply side, we use an agent-related approach to depict the stylized facts of the diffusion dynamics. The demand side is not dealt with in an agent-based manner, but rather by making plausible assumptions about its dynamics. Deriving aggregate effects of novelties for the whole economy necessitates distinguishing between these different levels of economic dynamics as well as relating them to each other.[7] This will be done by way of computer simulations (Section 4). Conclusions are drawn in Section 5.

2 Agent-Based Modelling of Behavioral Innovation Triggers

2.1 *Behavioral Analysis of Innovation Activity*

Under what conditions and in what way do agents create novelties or—in knowledge related terms—under what conditions do agents search for new knowledge? Setting aside a principal methodological caveat against this question,[8] there are two types of answers to it. In the 'functional approach' (mainly originated in the work of Hayek), a strategic (first mover) advantage for successful creators of novelties is derived from competition. From this assertion, it is directly concluded that there is a person/an agent who makes use of this advantage. In the 'personal approach' (mainly originated in the work of Schumpeter), it is assumed that there is a specific type of agent whose main profession is to innovate, i.e. the entrepreneurs. Both approaches are not sufficient in explanatory terms. The personal approach neglects the fact that innovating is a temporary activity which principally can be attributed to every economic agent. The functional approach fails to explain why, in a given time span, only a part of a whole population linked by a competitive process is in fact innovating and what kind of motives these innovating agents have.

The reason for this explanatory gap might lay in the above mentioned methodological caveat against the possibility of a microanalysis of novelty creation, which follows from the paradox that the unknown consequences of novelty creation make it difficult or impossible to analyze such a process. However, this gap of explanation exists only as long as we regard the material or immaterial outcome of innovative processes. If, however, we conceptualize novelty creation as a specific

[7] This procedure manifests the importance of the multi-scale property for analyzing the economy as a 'complex adaptive system' (Arthur et al. 1997).

[8] According to this caveat, the novelty creating process is totally conjectural without anything to generalize. Due to the idiosyncratic nature of the processes, as well as of the persons involved in innovations, some authors see only a limited possibility for an after-the-fact analysis on an aggregated level (Vromen 2001). For a critical discussion of these assumptions cf. Beckenbach and Daskalakis (2003 pp. 3)

way to act (having a contingent outcome), there are possibilities to analyze the determinants of the selection of this type of action. Such analyses might help to get a better understanding of innovation.

An approach of this type is confronted with the problem that evolutionary economics does not yet offer a satisfactory theoretical framework for it (cf. Endres and Woods 2010). The Carnegie School, however, does provide such a perspective on entrepreneurial acting. In the present context, the Carnegie approach is important in (at least) three respects: *First*, the approach sheds light on the determinants of the selection between the two relevant modes of action: routine acting (as the default mode of action) and search (in terms of shifting from persistence to change through novelty creation). *Second*, the decision units (especially firms) are not conceptualized as being internally consistent, but rather are seen as entities with internal conflicts being moderated in different ways and reaching common goals (Cyert and March 1992, p. 162, pp 229).[9] *Third*, a common denominator of the features of a decision unit is its limited ability to perceive information, and to transform this information into activity. This gives some hint as to how to specify novelty creating activities in behavioral terms.

According to the Carnegie School, the basic action mode of firms is routine acting. This action mode is continued at least as long as the degree of goal attainment is compatible with the aspiration level of the firm (March and Simon 1993, p. 205). Thereby it is supposed that the dimension of the firm's goals (i.e. in terms of production, inventory, sales, market share and profit (Cyert and March 1992, pp. 46–49)) are formed in complex and conflict-laden processes between individuals and subgroups of the firm (Cyert and March 1992, pp. 33). The aspiration level is set up as a weighted mean of the following three variables: "the organization's past goal, the organization's past performance and the past performance of other 'comparable' organizations" (Cyert and March 1992, p. 162). These aspiration levels are changing over time according to the current level of goal attainment and the perceived competition (March and Simon 1993, p. 204; pp. 33; March 1994, p. 29), leading to an adaptation of goals in its different dimensions (Cyert and March 1992, p. 172).

The comparison between the degree of goal attainment and the aspiration level (i.e. the degree of satisficing) is a core topic of the Carnegie School because it is supposed that the non-attainment of goals is triggering distinct entrepreneurial decision processes within the firm. In this context, we have to distinguish between two types of decision situations: on one side, there are situations in which a decision is made out of a given number of known alternatives; on the other side, there are situations characterized by the fact that the specific properties of possible alternatives are not known ex ante (March and Simon 1993, p. 195). The latter type of decision situation is a characteristic feature of innovation proper and is the focus of our analysis.

Following the assertions of the Carnegie School, innovations as search for new action alternatives can be triggered by falling short of as well as by exceeding the

[9] This focus allows for deriving characteristics on the firm level from characteristics at the individual level.

aspiration level ("satisficing search"; March 1994, p. 31; March and Simon 1993, p. 205). Here it is supposed that—due to its urgency—an innovation triggered by falling short of the aspiration level ("failure-induced search"; cf. Cyert and March 1992, pp. 188; March and Simon 1993, pp. 203) is more goal-directed and efficiency-oriented than a "success-induced search" (March 1994, 34) which is triggered by exceeding the aspiration level. The latter type of search requires the existence and availability of free resources ("slack") within the firm (which so far had only been used inefficiently) (March 1994, p. 34; Cyer and March 1992, pp. 188; March and Simon 1993, pp. 203; Levinthal and March 1982, p. 189). Due to its surplus in terms of resources, the success-induced search (also called slack search) is less restricted and therefore allows for more radical innovations than the failure-induced search (Levinthal and March 1982, p. 189; March 1994, p. 34).

However, falling short of or exceeding the aspiration level does not necessarily induce a search for new action alternatives. An adaptation of the aspiration level (reduction in the former case, increase in the latter case) can lead to persistence of the dominant action mode of routine.[10] One possible reason is that innovations are considered as sunk costs which can be avoided by continuing the routine mode (March and Simon 1993, p. 194; Simon 1997, p. 89). So the question arises which factors are crucial for firms' switching into search mode instead of sticking to a routine, even in case of not attaining or exceeding the aspiration level?

One such factor is risk acceptance (cf. March 1991a, p. 101; 1994, pp. 35–55).[11] Since the outcome of search processes is per se unpredictable and since search involves some costs, the risk acceptance of firms (resp. of decision makers within firms) determines the willingness to initiate search processes. Two aspects are relevant in this regard: Risk acceptance depends on individual factors and on the (perceived) situation. (March 1994, pp. 42) distinguishes between three classes of situations, depending on the degree of aspiration fulfillment: If the goal attainment exceeds the aspiration level or falls short of it, the risk acceptance is relatively high. If the aspiration level is met approximately, risk acceptance is lower, i.e., decision makers choose less risky strategies.

Corresponding to these different risk attitudes, it is possible to distinguish between more and less risky search strategies. Assuming that search processes for innovations can be conceived as problem solving processes (March and Simon 1993, pp. 197), a distinction can be made between two types of such problem solving: productive and reproductive problem solving (cf. March and Simon 1993, p. 198). The former involves the creation of fundamentally new options ('exploration'); the latter aims at processing options that are more or less existing, but not yet used ('exploitation') (March 1991b, p. 71; 1999, p. 184; 1994, p. 273; 2006, p. 205). Whereas exploitation is characterized by rather general features such as "refinement, choice, production [and] efficiency" (March 1994, p. 273), explorative

[10] Exceeding the aspiration level can also lead to a reduction of search activities of a firm (cf. March 1994, 31; Cyert and March 1992, pp. 41; March and Simon 1993, p. 194).

[11] For a discussion of the risk conceptions of managers, see March and Shapira (1987).

processes are assigned the above mentioned risk acceptance as well as more features pointing at a special (creative) propensity to discover something new: "experimentation, play, flexibility [and] discovery..." (March 1994, p. 273).

It becomes apparent that search processes, especially exploration resp. productive problem solving processes, require a high level of organizational capacity and skill of the members of the firm. With regard to the restriction imposed on the success of a problem solving process by bounded rationality (i.e. in terms of available knowledge), the Carnegie School also emphasizes the relevance of group problem solving (cf. March and Simon 1993, p. 173, pp. 201). Group problem solving allows for saving time by subdividing parts of the problem, for gaining access to different skills, for reducing error through communication, and for enhancing the quality of the solution (cf. March and Simon 1993, p. 202, p. 215; cf. also Marengo et al. 2000 Okada and Simon 1997).

However, the ability of a single firm to conduct group problem solving is restricted by time and skills. Furthermore, the "knowledge explosion" (March 1999, p. 181) which characterizes modern economies constantly increases the amount of (new) knowledge necessary for survival. Therefore, organizations have to consider whether they pursue the search endeavor alone or set up a (temporary) partnership with other firms. Unfortunately, the Carnegie School has little to say about this topic, as it predominantly deals with processes within organizations. But the huge body of research conducted to analyze innovation networks[12] gives evidence for the relevance of the topic. From a problem solving point of view, such cooperation might also be seen as a form of group problem solving (Daskalakis and Kauffeld 2007, p. 7; von Hippel 1990). Such cooperation constitutes a win-win situation for firms, as long as the players are competent, trust each other (Klimoski and Karol 1976; Noteboom 2004, p. 100; McEviliy and Zaheer 2006), and behave reciprocally by displaying relevant knowledge (Daskalakis and Kauffeld 2007, pp. 9, 17), and as long as the transaction costs required for setting up and maintaining the cooperation do not exceed the (presumed) gains (Beckenbach et al. 2009, p. 82; Lorenz 1999). It can be theoretically expected and empirically observed that firms exhibit heterogeneous attitudes towards engaging in a cooperation (Beckenbach et al. 2009, p. 85; Daskalakis and Kauffeld 2007, p. 14–19).

If a firm is involved in a process of group problem solving, three conditions have to be considered: *First*, there has to be a basis for communicating in terms of a mutual stock of knowledge being part of each involved firm; *second*, as the purpose of group problem solving lies in the exploitation of different skills, cognitive distance or complementary cognitive endowment as proposed by Noteboom (Noteboom 1992, p. 297; 2004, pp. 21; Noteboom et al. 2007) might be required; and *third* the knowledge must have the attribute of being transferable.

To sum up, the aspiration level plays an important role for the initiation of decision processes concerning the choice between continuing the current action

[12] We use the term in a broad sense, including the notions of cooperation, national and regional innovation networks, clusters, industrial districts, and so on.

mode (routine acting) and the introduction of a new action mode, namely innovation. With regard to the differentiation between success-induced search and failure-induced search as well as between productive and reproductive problem solving, at least two degrees of novelty inherent in innovation can be distinguished. The success-induced driver for search and the corresponding productive problem solving aim at a high degree of novelty and the introduction of new products. The failure-induced driver for search and the corresponding reproductive problem solving are primarily focused on improving existing products (of the firm itself) or on imitating products of the firm's competitors. Individual characteristics such as risk acceptance and other specific factors (which can be grasped by the notion of "exploration drive") play a crucial role in the decision on the degree of novelty to be created. Finally, the degree of complexity of modern problem solving processes requires a choice as to whether the search process is to be conducted alone or in cooperation with other firms.

2.2 Architecture of the Agents in the Simulation Model

2.2.1 Individual Modes of Action and their Triggering Conditions

The behavioral background features for innovation activities of firms emphasized by the Carnegie school are incorporated in the present modeling of agents. It is assumed that the firm-agents pursue two goals, profit and market share, for each of which an aspiration level is set. According to the above mentioned, insights the firms are able to perform three modes of action: routine behavior (as default mode), imitation and innovation. Whether an agent leaves the default mode of routine and what kind of novelty creation is selected, depends on different feasibility conditions and on behavioral forces composed of indicators for the goal attainment (in terms of the aspiration level), expected cost for the target activity, and the amount of slack.

Before any kind of novelty creation can be pursued by the firm, the financial feasibility has to be checked. This is done by comparing the current liquidity and the profit expected from the novelty creation[13] on one side, and the expected costs of novelty creation on the other side.[14] If the latter are covered, the novelty creation is classified as feasible.[15]

[13] The expected profit is calculated by a linear regression on past profits weighted by a parameter reflecting angent-specific degrees of optimism/pessimism.

[14] The expected costs are calculated by multiplying the given costs per time step and the average time for novelty creating processes. According to empirical findings, it is assumed that the costs of an imitation project are lower than the costs of an innovation project.

[15] If the novelty creating process is intended as an imitation, the knowledge about another firm's product (improvement) which appeared recently in the market is required as an additional feasibility condition.

According to the concept of the Carnegie-school sketched above, the basic driver in favor of novelty creation is assumed to be the relationship between goal attainment in terms of profits (p) and market share (m) and the respective aspiration level (asp, asm). In formal terms, this means[16]

$$f_1(t) = w_1 \left(\frac{asp(t)}{p(t)}\right)^{\varepsilon_1} \tag{1}$$

$$f_2(t) = w_2 \left(\frac{asm(t)}{m(t)}\right)^{\varepsilon_2}. \tag{2}$$

In each time step, the aspiration levels are updated according to:

$$asp(t+1) = (1-\varphi)\, asp(t) + \varphi\, p(t) \tag{3a}$$

$$asm(t+1) = (1-\varphi)\, asm(t) + \varphi\, m(t) \tag{3b}$$

where ϕ is the flexibility of adaptation, which is another individual trait of the firms ($0 \leq \phi \leq 1$) reflecting the relative weight attached to past and present performance. The force in favor of *failure induced search* (corresponding to imitation) can be formalized then by relating the sum of f_1 and f_2 to the expected costs of an imitation project:

$$F_1(t) = \frac{f_1(t) + f_2(t)}{cim}. \tag{4}$$

In the case of *success induced search* (corresponding to innovation), f_1 and f_2 will be low. Specifying the triggering conditions for innovation, therefore, necessitates take into account other factors such as slack, risk and the willingness to explore. Considering reserves in terms of financial resources (fr) and in terms of knowledge (kr)[17] as the main sources for slack, and conceptualizing the exploration drive (w_0) as a weighting parameter for this slack, the component in favor of innovation can be formalized as:

$$f_0(t) = w_0\, (kr(t) + fr(t)). \tag{5}$$

Summing up this additional component and the factors for failure induced search (which still have a weak influence) and relating this sum to the expected cost of an innovation project is tantamount to the innovation driver. Integrating the risk

---

[16] For reasons of model fine tuning, these relationships are formalized as elasticities (power of ε) with different weights (w).

[17] The knowledge reserves (kr) are operationalized as the relation of the number of sharable knowledge domains (see below) of the agent to the total number of sharable knowledge domains; the financial resources (fr) are operationalized as the share of the current profit in relation to current turnover.

attitude (α) as a weight for this driver gives the formal expression for the forces in favor of innovation:

$$F_2(t) = \alpha \frac{f_0(t) + f_1(t) + f_2(t)}{cin}. \tag{6}$$

Taking the routine mode as the default mode of activity (to overcome by novelty creation)

$$F_0(t) = 1 \tag{7}$$

as a reference value,[18] the actual mode of action pursued by the firm (F_{am}) is determined according to the condition:

$$F_{am}(t) = \max(F_0(t), F_1(t), F_2(t)). \tag{8}$$

To allow for heterogeneity in the population with regard to the behavioral parameters the firms are subdivided into three different types: *first*, conservative firms, being more aversive against changes, hence fostering routines; *second*, cautious firms, which prefer imitation; and *third*, experimental firms, which are more likely to pursue innovations. The firms differ with respect to the behavioral parameters α, w_0, ε and ϕ. The distribution of the type of firms as well as the values of α and $w_{0...2}$ are derived from an empirical investigation of the novelty creating behavior of German firms (Beckenbach et al. 2009).

2.2.2 Cooperative Innovation

A firm agent who has selected the action mode of innovation still has two options (see above): He can try to develop an innovation on his own (individual innovation) or he can seek cooperation in order to enable or facilitate the development of an innovation.

Analogously to the selection mechanism between the three action modes (discussed in Section 2.2.1) and the selection between individual and cooperative, innovation is determined by a comparison of forces: Denoting the propensity to cooperate by χ and the share of cooperative innovations N_c (related to the total number of innovations N) by $\frac{N_c}{N}$, the cooperation force of an agent is given by:

$$cp(t) = (1 - ifb - iff)\,\chi + ifb\,\frac{N_c(t)}{N(t)} + iff\,\sum_{i=0}^{2} f_i(t), \tag{9}$$

[18] Setting the preservation force as a constant is no restriction of generality since the absolute values of the forces F_i do not matter; it is only the ratio between them which determines the action mode. There are three special or exceptional cases in which the selection mechanism mentioned above is not applied (or even not applicable) two of which concern start up firms and one concerns firms with a negative or zero profit. (See Beckenbach et al. 2009 for a more detailed discussion.)

ifb and iff being parameterized weights for the different triggering forces. Actually, a firm agent is willing to cooperate if this cooperation force is larger than a threshold (*ct*):

$$\mathrm{ct} < \mathrm{cp(t)}. \tag{10}$$

Given the willingness of an agent to cooperate, a matching procedure takes place to determine if a cooperation can actually be formed. The *matching conditions* concern the existing relationship between firm agents, as well as some knowledge requirements. As to the relationship between firm agents, it is required that the partners for a cooperation should either belong to the same branch or that they be related to each other as suppliers or customers.[19] Moreover, potential partners to whom an agent has a high level of trust are preferred because trust facilitates the exchange of knowledge and reduces transaction cost. As to the knowledge requirements, there is a condition that requires a certain amount of common (non-rivalrous) knowledge (accessible at low costs), ensuring a common basis to communicate within the cooperation process, and another condition that requires a certain amount of complementary (specific) knowledge (accessible only at high costs), constituting an incentive for knowledge exchange. The overall matching of firms in the process of co-operative innovation is summarized in Fig. 1.

Once a cooperation has been formed, a process of knowledge transfer starts. This process lasts a fixed number of time steps (the development duration of an innovation), in each of which knowledge in one domain is transferred from one agent to the other with a certain probability. This probability (pb) depends, on the one hand, positively on the trust (tr) of the knowledge-giving agent in the knowledge-receiving agent and, on the other hand, positively on the absorptive capacity (*ac*) of the receiving agent.[20] Formally, the probability is given by:[21]

$$\mathrm{pb(t)} = se\ (\mathrm{tr(t)} - 1) + ac, \tag{11}$$

where *se* denotes the sensitivity of the probability of transferring knowledge with respect to trust.

Each time this knowledge transfer does not happen, the trust of the second agent in the first is diminished by a certain decrement; by contrast, trust is raised by a certain increment each time the transfer actually happens. If trust falls beyond a certain threshold, the second agent breaks off the cooperation and abandons this innovation project. If there remains more than one agent from the current

[19] At the beginning, each firm is assigned randomly a set of supplier firms, being fixed for the whole simulation.

[20] The absorptive capacity is conceptualized as a given probability weight for the cooperation to happen.

[21] If the right hand side of the equation is negative, the probability is set to 0. We assume $0 \leq \mathrm{tr} \leq 1$ and $0 \leq ac \leq 1$.

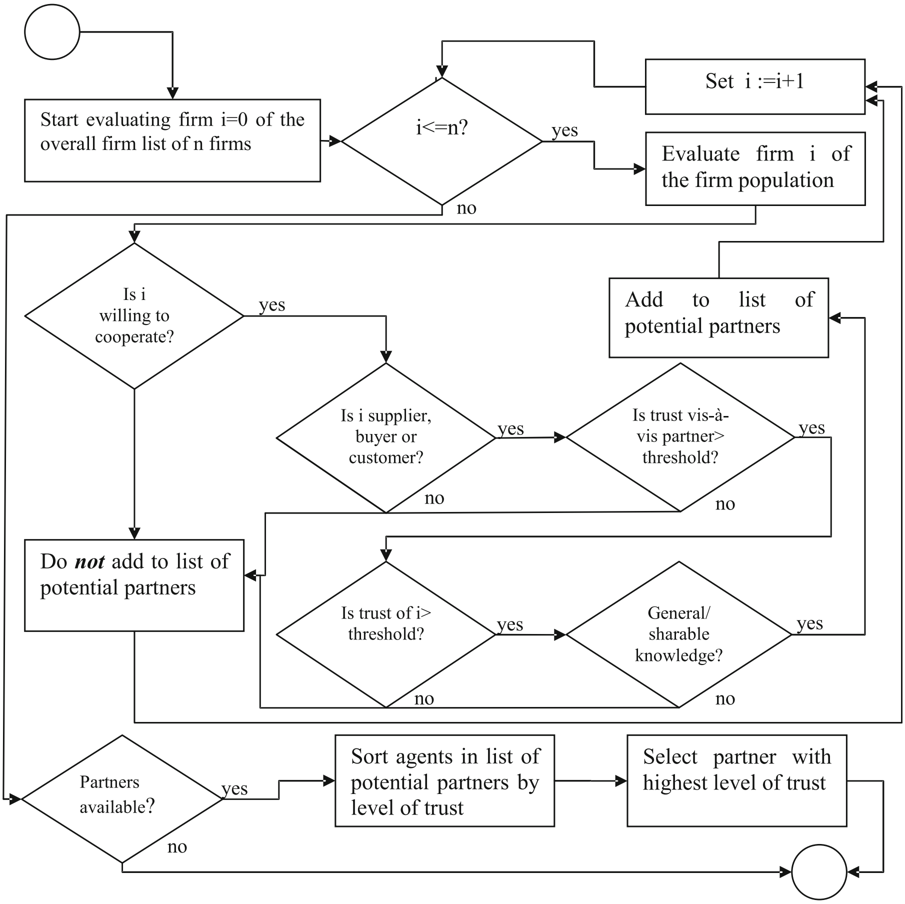


Fig. 1 Matching of firms in co-operative innovation

cooperation and if these agents have successfully exchanged some knowledge, they continue the cooperation. The decrement and increment of trust as well as the threshold are model parameters. Each step of this transfer process has a certain cost for both partners.

If the transfer process is successfully accomplished, an innovative product is created and put on the market, as described above. Production costs and returns are shared equally among all partners.[22]

[22] Independently of the knowledge transfer described above, there are two probabilistic "knowledge destruction" processes taking place in each time step: a decay (forgetting) of knowledge for each firm and a general devaluation (depreciation) of knowledge for all firms as a global effect of technological change.

3 Dynamics of Diffusion on the Sectoral Level

3.1 Conceptual Foundation

Diffusion is the process by which innovations are adopted by users within a population of agents (Jaffe et al. 2003). In this regard, a basic but important observation is that diffusion is time consuming. Moreover, the diffusion process typically produces a characteristic S-shaped curve if the number of adopters is plotted against time (Dosi 1991; Rogers 1995). In the empirical literature, several other factors have been analyzed that explain both scope and speed of innovation diffusion (Griliches 1957; Mansfield 1968; Gort and Klepper 1982; Rogers 1995; Grübler et al. 1999). Insights from this literature include (among others) the observation that diffusion hinges on the aspects of the underlying technology (both the innovative technology and the technology to be substituted), environmental incentives, the characteristics of adopters, as well as the spreading of information (Dosi 1991).

In the modelling of diffusion at least three influential approaches can be discerned (cf. for surveys Metcalfe 1988; Dosi 1991; Kemp 1997; Geroski 2000; Jaffe et al. 2003). *First*, epidemic models, which consider the diffusion process in analogy with the spreading of an infectious disease. This modelling approach particularly stresses the importance of the spreading of information as a prerequisite for adoption. The model itself was mainly criticized for not delivering (micro-founded) explanations for the adoption of innovations (Dosi 1991).[23] *Second*, there are so-called probit models which explicitly take into account the adoption decisions of the agent—usually in the form of choice based on optimizing behavior (Kemp 1997, p. 80). Hence, in these models, aspects of learning and interaction within the population of agents do not play a prominent role. A *third* approach is given by agent-based models of innovation diffusion (Silverberg et al. 1988, Windrum and Birchenhall 2005, Schwoon 2006, Schwarz and Ernst 2009, Wittmann 2008, Windrum and Birchenhall 2009a, 2009bb). In these models—mostly related to products with different environmental qualities—the interaction and decisions of heterogeneous agents are an integral part of the modelling concept. Considering the criticism regarding the other categories, agent-based models have the advantage of capturing both the heterogeneity of the potential adopters as well as the particularities of various forms of interaction (including knowledge transfer). For this reason, they can provide micro level details for the analysis of innovation diffusion.

Apart from the spreading dynamics for the innovation itself, the diffusion dynamics is strongly influenced by the conditions for exit and entry in a given sector. The intensity of competition, on one side, and the financial constraints, on

[23] Cf., however, Geroski (2000) for an exposition of how a basic diffusion model can be enriched to incorporate heterogeneity on the part of the adopting agents.

the other side, are considered to be most influential in this respect (Dosi et al. 1995; Agarwal and Gort 1996).

Because the main focus of our agent-based perspective is on explaining the occurrence of novelty creations in different forms, the explanation of the diffusion dynamics suggested here combines stylized facts of diffusion research and agent-based analysis: A successful novelty creation on the individual level (in terms of product improvement) triggers a demand dynamic for this product. If initially a critical mass for this product is overcome, a self-feeding diffusion process up to a maximum level in which all market potential is exhausted takes place. According to the dominance of retarding effects at the beginning of this diffusion process, and due to the dominance of the promoting effects at later stages, an s-shaped time-dependent diffusion curve is assumed (Rogers 1995). Finally, the twofold nature of economic activities (see Section 1) necessitates giving innovations a twofold sectoral effect: a growing of the demand for the products of an innovating firm and a substitution effect for older products. Hence, the overall diffusion process is not simply postulated or constructed but is strongly shaped by the type and the time in which the novelty creation is occurring on the individual level.

3.2 Features of the Diffusion Process in the Simulation Model

The starting point of the diffusion dynamics is the emergence of an innovation (developed according to the agent-based explanation in Section 2.2). As innovations are high risk ventures, we assume that there are two possible outcomes of a diffusion process:

- on the one hand, there are certain innovations successfully diffusing in a population by successively increasing their share of demand;
- on the other hand, there are innovative failures which may temporarily attract some demand, but finally fail to diffuse on a large scale.

Formally, we discern between these two outcomes by calculating a demand potential (ypo) that represents the total demand possible for an individual innovation. The demand potential functionally depends on the current turnover of the innovator (u) as well as the amount of knowledge with which the innovating agent is endowed (k). The assumption behind the last point is that a broader knowledge base is correlated with a higher flexibility as well as a better ability to meet the demand. This superior knowledge-dependent ability of an agent is assumed as giving access to a higher demand potential (ypo) for the agent's products:

$$\mathrm{ypo(t) = u(t)k(t)}\,\lambda_1 \tag{12}$$

where λ_1 is a stochastic term depicting market uncertainties. In our model, *successful innovations* are those for which the initial demand potential exceeds a threshold value yts. We assume that

$$\mathrm{yts(t)} = \mathrm{ypo(t)}\lambda_2 \tag{13}$$

where λ_2 again is a stochastic term depicting market uncertainties. Given this constellation, the demand gradually approaches the demand potential as the simulated time progresses. A parameter v regulates the speed of this process for successful innovations, which can be formalized as:

$$\mathrm{y(t+1)} = \mathrm{y(t)} + v\frac{\mathrm{(y(t) - yts(t))\ (ypo(t) - y(t))}}{\mathrm{ypo(t) - yts(t)}} \text{ if } \mathrm{y(t)} \geq \mathrm{yts(t)}, \tag{14}$$

$$\lim_{t\to\infty} \mathrm{y(t)} = \mathrm{ypo(t)} \text{ if } \mathrm{y(t_0)} > \mathrm{yts(t)} \text{ and}$$
$$\lim_{t\to\infty} \mathrm{y(t)} = \mathrm{yts(t)} \text{ if } \mathrm{y(t_0)} = \mathrm{yts(t)}.$$

The second type of innovations, *innovation failures*, are those in which the condition involving the threshold condition is not met, i.e. where the relevant threshold value is not reached at the time the innovation enters the market. In this case, initial demand successively diminishes during the following steps until it converges to 0.

$$\mathrm{y(t+1)} = \mathrm{y(t)} + v\ \frac{\mathrm{y(t)\ (y(t) - yts(t))}}{\mathrm{yts(t)}} \text{ if } \mathrm{y(t)} < \mathrm{yts(t)} \tag{15}$$

$$\lim_{t\to\infty} \mathrm{y(t)} = 0 \text{ if } \mathrm{y(t_0)} < \mathrm{yts(t)}.$$ [24]

Both innovations and imitations are crucial features of environments characterized by competition. In our model we assume that successful imitators can attract a certain share of the demand potential created by an innovator. Technically, this means that intertemporal increases of demand (based on the diffusion dynamics described above) are equally shared among innovators and imitators of a specific product if $\mathrm{y(t)} < \mathrm{ypo(t)}$.

Another remarkable characteristic of the diffusion of innovation is that it replaces conventional products to a certain extent. This can be thought of as a form of creative destruction which leads to a situation in which innovators grow at the expense of other firms that lose part of the current demand. On the other hand, it is possible that innovations are not directly competing with conventional products and create additional demand. These considerations are taken up in our model by introducing a substitution parameter *su* that measures the degree by which conventional products are replaced by innovative products.

[24] An intermediate outcome emerges if $\mathrm{y(t_0)} = \mathrm{yts(t)}$. In this case, the demand approaches the threshold value yts(t).

Referring again to the diffusion dynamics, we can formalize the growth of demand due to innovative products for a certain sector (where W denotes the absolute growth of demand and k={1..r} are the innovations in the sector under consideration):

$$\mathrm{W}(\mathrm{t}+1)=\sum_{k=1}^{r}(\mathrm{y_k}(\mathrm{t}+1)-\mathrm{y_k}(\mathrm{t})). \tag{16}$$

But, as new demand substitutes for conventional demand to a certain degree, we can define, with the help of the substitution factor, an overall growth rate of demand. Setting $\mathrm{Y}(\mathrm{t})=\sum_{k=1}^{s}\mathrm{y_k}(\mathrm{t})$ (s denoting the number of agents in a given sector) it follows:

$$\frac{\mathrm{Y}(\mathrm{t}+1)-\mathrm{Y}(\mathrm{t})}{\mathrm{Y}(\mathrm{t})}=\frac{(1-su)\,\mathrm{W}(\mathrm{t}+1)}{\mathrm{Y}(\mathrm{t})}. \tag{17}$$

For the extreme cases of the substitution factor, this means:

- if $su=1$, all additional demand of innovations replaces current demand. There are no net growth effects,
- if $su=0$, all additional demand from innovations is added to the existing demand.

For intermediate values innovation can be thought of as showing both characteristics: it replaces some part of conventional demand and adds a new component to it.[25]

All the aspects considered so far are of crucial importance for the financial performance of the firm. This is caused by the fact that demand for its products generates a feedback to the financial performance of the firm. In particular, we introduce the following features in our modelling framework:

- for each product, there exists a linear return and a linear cost function which depend on demand as described by the difference equations above;
- there is an additional cost component associated with innovation projects (this is a model parameter that can be varied);
- profit is turnover originating from sales of innovative and conventional products (depending on the return function) minus variable and fixed cost.

Financial resources are updated according to profits (losses) of the current period. The financial performance of the firm can finally give rise to another reason a product disappears from the market. It is assumed that a firm stops producing a

[25] In order to achieve this sectoral outcome, the demand facing the individual firm is rescaled by a scaling factor: $sf(t)=\frac{Y(t)+(1-su)\,W(t+1)}{Y(t)+W(t+1)}$.

certain product once the revenues of a product no longer cover production costs. This happens when the demand for such a conventional product diminishes, especially by the substitution mechanism mentioned above.

The *entry* of new firm agents is modelled probabilistically: In each time step and for each production sector, a new firm agent is created with a probability that depends linearly on the sector-specific indicator of competition intensity and on the amount of public subsidies for new entrants. Whereas the indicator of the competition intensity is subject to an exogenous linear dynamic, the amount of public subsidies is a model parameter. An entrant firm does not sell any conventional products; it always starts with an attempt to create an innovative product (either individually or cooperatively). The behavioral type of an entrant firm agent is set probabilistically as well; the probability distribution corresponds to the shares of the behavioral types given at the beginning of the simulation when the initial firm agent population was created.

If the financial resources of a firm agent fall beyond zero (meaning the agent runs into debt), the agent *exits* the market. This can happen in the course of an innovation or imitation project if the actual profit (gained by conventional products and innovative products that already have been put on the market before) falls behind the expected profit (which the agent calculates on the basis of the profit in the previous time steps), or if the actual development duration and, consequently, the actual development costs exceed the expected duration or costs, respectively.

To sum up: This modelling of the diffusion process is a combination of agent-based and parametrized components. The agent-based component determines the point in time as well as the type of the occurring novelty. Furthermore, according to the performance and the knowledge endowment of the innovating agents, a dynamic market potential as well as critical mass is defined for every product. Based on that, the distribution of the first mover advantage between novelty creating agents is determined exogenously. What is beyond the scope of this analysis is a fully fledged modelling of the adopters of new products, in that especially the velocity of market potential exhaustion and the amount of substituting old products by new one are derived from an explicit modelling of adopter agents. These effects are parametrized in the diffusion dynamics presented here.

Both aspects of innovation, namely creation and destruction, can be found at the level of diffusion as well. This includes, on the one hand, the creation of additional demand by innovative products. On the other hand, there is an inherent dynamics of demand destruction for conventional products which is particularly captured by the substitution factor. Additionally, from the point of view of an individual firm, innovation-induced destruction emerges if a part of its demand is competed away by imitators as well as in the form of a decision to abandon a product delivery once it is no longer viable in terms of covering the production costs.

4 Simulation Results

4.1 Exemplary Analysis of a Singular Run

To explore the emergence of sectoral growth and the impact of heterogeneity of firm agents, we show results of a simulation run with four production sectors and three types of firms being different in behavioural terms. Each production sector comprises 23 firm agents, among which seven belong to the conservative behavioral type, five to the cautious and eleven to the experimental type.[26] Table 1 summarizes the values of the most important model parameters for the standard configuration.

According to the architecture of the simulation model, the economic outcome can be analyzed on different levels of aggregation. At the highest level of aggregation (the 'national' level) is aggregated net value (i.e. the profits over all sectors) over time. The parameter settings summarized in Table 1 are tantamount to an optimistic scenario, leading to an continuous increase of net production in 120 time steps (30 years) (Fig. 2; top). Below this aggregated level there is an uneven development of the different sectors due to their internal dynamics in terms of novelty creation (Fig. 2; bottom). As regards the action modes, it can be observed that, after the transient phase (up to about $t = 38$; cf. dotted lines in Figs. 2(top) and 3), an inverse cyclical development of innovation and imitation frequencies can be observed, indicating that phases dominated by innovation and phases dominated by imitation alternate. In this quasi-regular state in terms of action modes, the growth of the net value increases significantly.

In the given simulation framework, the generation of new knowledge is not an exogenous element; it is strongly tied to the innovation process. According to the slowly increasing share of innovation activities (Fig. 3), the societal endowment with knowledge increases moderately. Figure 4 (top) shows this moderate increase of the overall level of knowledge in terms of the average number of domains, about which the agents have knowledge. Disaggregating this indicator to the level of sectors manifests different sectoral knowledge dynamics (Fig. 4; bottom). The impression that novelty creation produces increasing heterogeneity in terms of knowledge is backed by further disaggregating the societal knowledge stock in different domains (Fig. 5; top) and by taking into consideration distinct modes of knowledge acquisition (Fig. 5; bottom). Disaggregating the knowledge stock into agent-specific components (sum total of individually disposable knowledge components) and tracing these components in the course of time shows, for many agents, an increase of knowledge and, for some agents, a decrease resulting in a persistent heterogeneity between individuals (Fig. 6; left); the lighter the grey, the higher the knowledge—white representing 'no agent existing', i.e. exit and late

[26] These shares for the behavioral types are taken from findings of an empirical investigation in the region of Northern Hesse in Germany (Beckenbach et al. 2009).

Table 1 Important model parameter settings

Parameter	Value	Parameter	Value
Mean duration of innovation	6	Cooperation threshold (*ct*)	.5
Mean duration of imitation	3	Absorptive capacity (*ac*)	1
Cost of innovation per time step	.125	Sensitivity of knowledge transfer probability w.r.t. trust (*se*)	1
Cost of imitation per time step	.0625	Initial value of trust	.75*
Exploration drive (w_0)	.15; .2; .25*	Trust increment	.1
Risk acceptance (α)	2; 2.5; 3*	Trust decrement	.2
Weight of profit aspiration (w_1)	0.04*		
Weight of market share asp.(w_0)	0.05; 0.04; 0.06*		
Mean flexibility of adaptation (ϕ)	.25	Trust threshold	.5
Propensity to cooperate (χ)	.7; 1; 1.3*	Substitution factor (*su*)	.75
Weights of components of coop. force (*ifb*, *iff*)	.25	Subsidies for an entrant	.5

If there are three values separated by semicolons in the "value" column, these values correspond to the three behavioral firm types (in the order conservative, cautious, experimental) according to the empirical findings mentioned above. The values are either based on these empirical findings (*) or on estimates

entry). Further disaggregation into the time-dependent knowledge of an arbitrarily chosen agent (Fig. 6; right) illustrates that the domain-specific knowledge is discontinuous, depending on forgetting and devaluating the existing knowledge (white representing areas of knowledge).

4.2 *Sensitivity Analysis*

It has been argued above that, from the perspective of innovation analysis, the economic activities have a dual nature: on one side they tend to replicate and, on the other side, there is an endogenous mechanism for disturbing this well-established state of affairs. Correspondingly, the innovation has two sides: a creative and a destructive one. Breaking routines, devaluating knowledge, replacing conventional demand and the exit of firms are different dimensions of the destructive side of innovation.

To demonstrate this two-sided effect of innovation, we select the parameter determining the replacement effect for conventional demand (*su*). In the first row of Fig. 7, it is shown how the outcome of the innovation dynamics in terms of total profit, frequency of innovation activities and the number of exits is changing if this substitution parameter is increased.[27] In all these outcome indicators, the

[27] Values in Fig. 7 are cumulated average values of the respective indicators at the end of the simulation run.

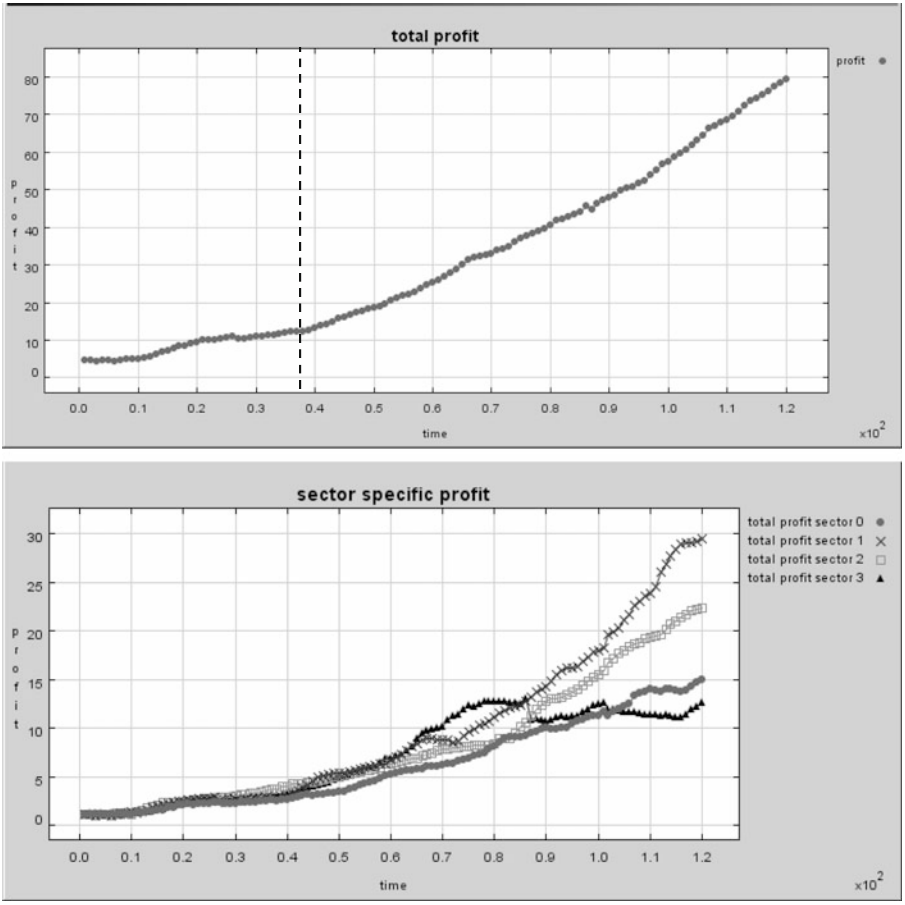


Fig. 2 Total profit (*up*) and sector specific profits (*down*)

destructive effect given by an increased ease to replace conventional demand (high values of *su*) can be verified. Due to intensified competition, the constraints for innovation become more critical and the risk of exit increases.

In the second row of Fig. 7, it is depicted to what degree an increase in the aspiration adaptability (more precisely, the flexibility to adapt the aspiration level related to the different goals symbolized by ϕ) can partly compensate for this destructive effect as regards the outcome indicators.[28] Although the frequency of innovation can be increased by increasing this adaptability, the positive overall effect on total profit is moderate due to demand constraints (at high levels of *su*).

[28] For each combination of values for these parameters, a Monte Carlo simulation with 16 runs with different seed values for the random number generator has been run. The darker the grey, the higher the respective indicator.

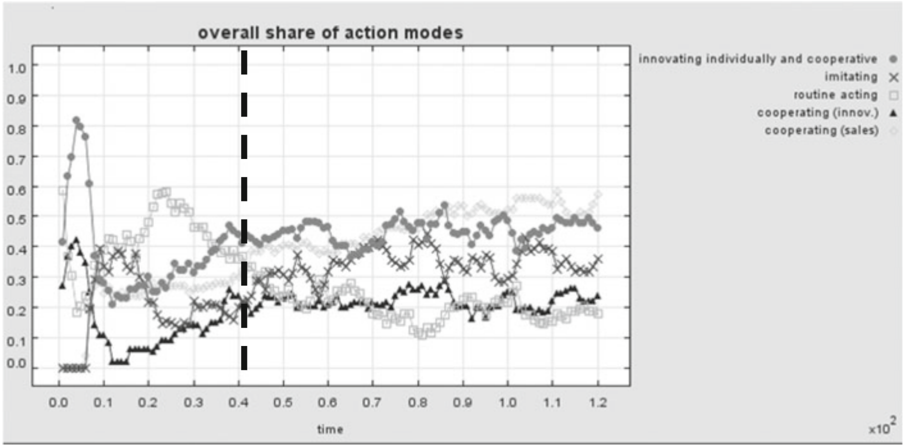


Fig. 3 Shares of action modes over time. The shares of innovating, imitating and routine acting agents sum up to 1 as each agent acts in each time step in exactly one of these three modes (they are mutually exclusive). However, agents participating in innovation cooperation form a subset of innovating agents. Hence, their share is never greater than the share of innovating agents. The participation in a sales cooperation is compatible with any action mode, so the share of agents participating in such a cooperation is independent of the shares of other action modes

The results regarding total profit are not systematically influenced by the impact of ϕ. Profits per innovation project, however, tend to be lower if ϕ is increased: innovators tackle more albeit less profitable innovation projects. Although the substitution parameter itself is positively correlated with the frequency of exits the adaptability has almost no influence on the former.

Looking at the third row of Fig. 7, one can conclude that a compensation for creative demand destruction is more likely, if the exploration drive (w_0) is increased: for every level of *su*, an increase in this drive to innovate is tantamount to an increase in the number of innovations. This increase of innovation frequency is sufficient for increasing the total profit for low and medium range levels of *su*. But in a more competitive environment (high values of *su*), innovation failure is more likely and cannot be offset by increasing exploratory activities (high w_0). In this case, the number of exits is not significantly reduced if the exploration drive is increased because the impact of competitive pressures (high level of *su*) becomes more important.

Finally, in the fourth line of Fig. 7, it is shown what happens if an increase in innovation induced substitution of demand is accompanied by an increase of the propensity to pursue cooperative innovation (χ). Not surprisingly, the frequency of innovation is increasing due to the facilitating effect of such a cooperation. But, as in the case of increased adaptability, the amount of additional innovation activities is not sufficient to have a significant effect on total profits. Finally, a higher number of exits is associated with a low level of cooperation drive (low values of χ) and more intense competition (high values of *su*). In this regard, a higher propensity towards cooperative innovations seems to safeguard against exits.

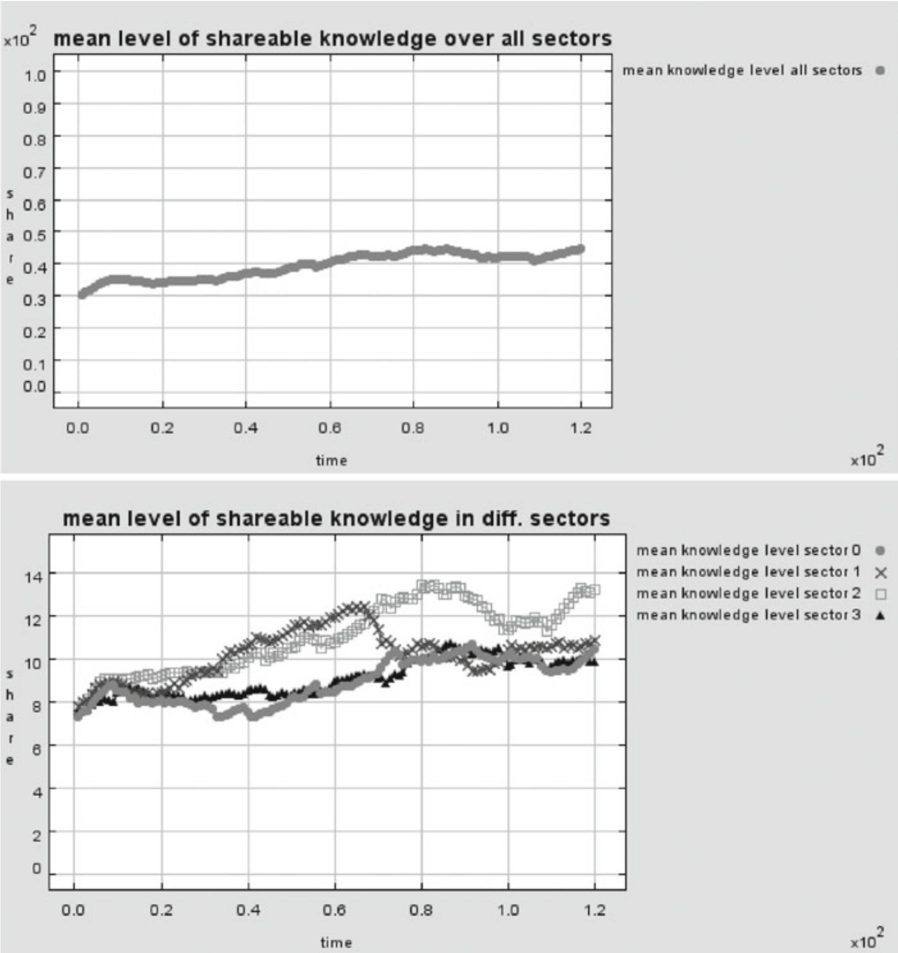


Fig. 4 Mean level of sharable knowledge over time

To sum up: if considered in isolation, behavioral parameters favoring innovation are appropriate for compensating the effect of creative demand destruction only in a limited way. For medium and high range values of the substitution parameter, a compensating boost of innovation dynamics is required which can only be achieved by an extraordinary combination of specific behavioral parameters changes. Hence, the normal state of affairs seems to be that innovations have a creative and destructive effect at the same time, the balancing of which determine the dynamics of sectoral growth.

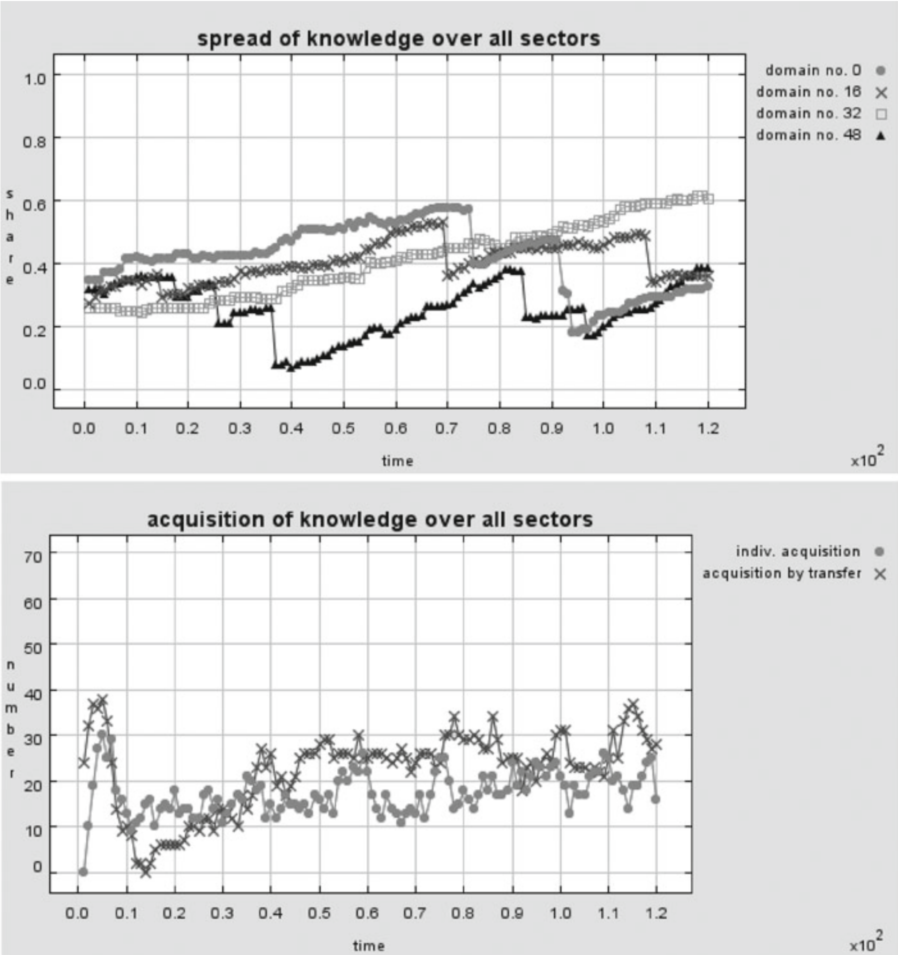


Fig. 5 Distribution and acquisition of sharable knowledge over time

5 Conclusions

One of the aims of the model suggested here is to specify the link between novelty creation and sectoral growth by including the internal trigger conditions for this type of economic activity. Explaining these trigger conditions necessitates a behavioral foundation for agents pursuing such a novelty creation. Not only situations favourable for the occurrence of novelty creating activities but also the particular form they come about are within the realm of such an explanation. Hence, a population of agents differing in their propensity to activate different modes of action can be derived and focused as the main explanans for the overall growth. It has been shown that such a behavioral foundation can draw upon the contributions made by the Carnegie School.

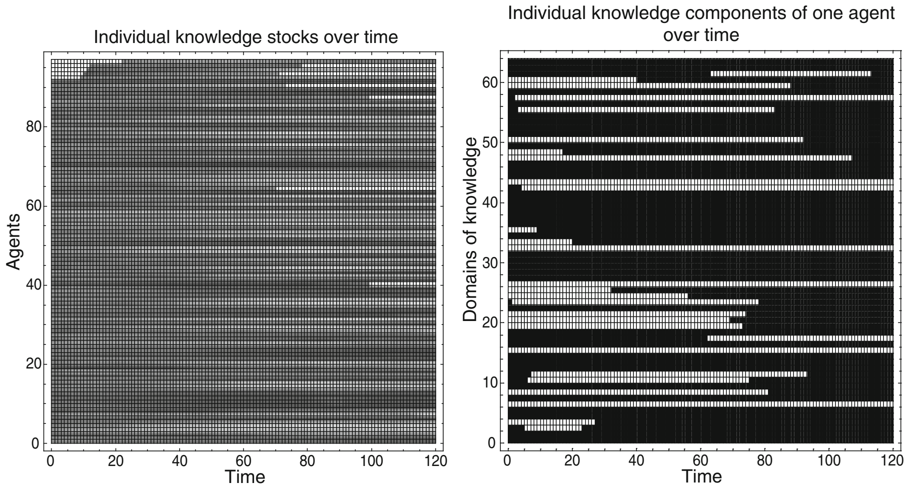


Fig. 6 Individual level of sharable knowledge over time

But to consider single agents triggering novelty creating activities is only the starting point for assessing the corresponding growth effect. The results of these activities have to be accepted by the demand and, if so, they are feeding an interactive diffusion process at the sectoral level. As a result of the different qualitative nature and of the different scale of the dynamics on these levels, a multi-level approach is an appropriate method for analyzing the sectoral growth dynamics.

A common denominator of these spreading effects of novelty creation is its twofold nature: apart from being 'creative' they are also 'destructive' in that established behaviors, artefacts and processes are overcome or degraded. On the behavioral level, established routines are broken, on the sectoral level conventional processes and products are replaced. Against this background, another aim of the model suggested here was to show how the creative and the destructive components of novelty creation are constituted on the different levels and how the balancing of these components creates an overall (growth) effect.

Finally, at least some of the well-known stylized facts of novelty creating processes can be reproduced by the suggested model:

- agents are not homogenized in the course of time; rather their heterogeneity in terms of knowledge and action modes persist (Dawid 2006, p 1241);
- the share of innovating agents is an emerging property and corresponds roughly to the statistical data (Rammer 2006);
- after the transient phase, there is a cyclical pattern correlating innovation and imitation inversely signifying that there are succeeding phases dominated by (radical) market innovation and phases with (incremental) imitation (Utterback 1996, pp 79).

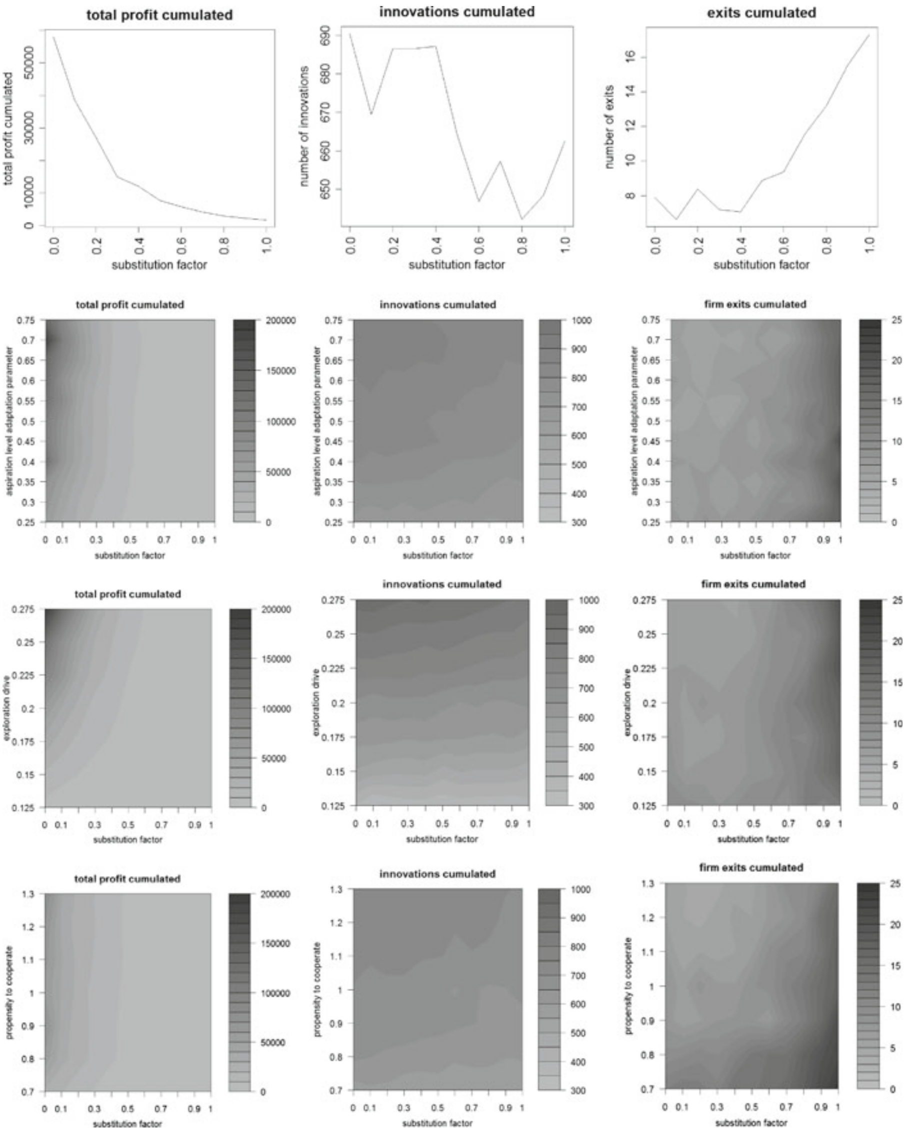


Fig. 7 Sensitivity analysis for the substitution parameter and various behavioral parameters

Further elaborating on a behaviorally enriched evolutionary growth model seems to be a promising path for future scientific research. Deepening the behavioral foundations by taking into account the composed nature of (at least medium and large sized) firms, endogenizing the demand side of the diffusion dynamics and specifying the technology-driven sectoral input/output coefficients as a basis for integrating the interdependencies between sectors will be some areas this research will have to tackle.

Appendix: Core of the Simulation Model (Pseudo-Code)

```
Determine component forces {
        Determine component force for slack-based search (cf. eq. 5)
        Determine component force for profit aspiration (cf. eq. 1)
        Determine component force for market share aspiration (cf. eq. 2)
}
Determine aggregate forces {
        Determine innovation force (cf. eq. 6)
        Determine imitation force (cf. eq. 4)
        Determine routine force (cf. eq. 7)
}
IF ((innovation force OR imitation force)>routine force){
        action mode := innovation mode
}ELSE{
        action mode := routine mode
}
IF (action mode=innovation mode){
         Shall perform knowledge acquisition := FALSE;
        IF (is agent currently involved in individual innovation project = TRUE OR
           is agent currently involved  in cooperative project = TRUE){
                Shall perform knowledge acquisition := TRUE;
        }ELSE{
                IF (willingness to cooperate < threshold value for cooperation (cf. eq. 9 and 10)){
                        Start new individual innovation project
                        Shall perform knowledge acquisition  :=  TRUE;
                        Type of knowledge acquisition := individual knowledge acquisition
                }ELSE IF (willingness to cooperate >= threshold value for cooperation (cf. eq. 9 and 10)){
                        Identifiy potential cooperation partners
                        IF (agent has potential cooperation partners  = TRUE){
                                Select cooperation partner
                                shall perform knowledge acquisition := TRUE;
                                type of knowledge acquisition := cooperative knowledge acquisition

                         }ELSE{
                                shall perform knowledge acquisition := FALSE;
                                type of knowledge acquisition := individual knowledge acquisition
                        }
                }
        }
}
IF (shall perform knowledge acquisition = TRUE){
        IF (type of knowledge acquisition = individual  knowledge acquisition){
        Try to acquire new knowledge individually
        }ELSE IF (type of knowledge acquisition = cooperative knowledge acquisition){{
                Try to acquire new knowledge from cooperation partner
        }
        IF(current time step = predefined end of innovation project){
                IF(is end of project reached = TRUE){
                        Terminate innovation project successfully
        }
                ELSE{
                Continue project next round
        }
        ELSE{
                Terminate innovation project unsuccessfully
        }
}
Determine turnover and profits
Update aspiration level (cf. eq. 3a/b)
Terminate unprofitable projects
```

References

Agarwal R, Gort M (1996) The evolution of markets, entry, exit and survival of firms. Rev Econ Stat 15:489–498

Arthur B, Durlauf SN, Lane DA (eds) (1997) The economy as an evolving complex system II. Addison-Wesley Publishing Company, Reading/MA

Beckenbach F, Daskalakis M (2003) Invention and innovation as creative problem solving activities - a contribution to evolutionary microeconomics. Volkswirtschaftliche Diskussionsbeiträge. http://www.wirtschaft.uni-kassel.de. Accessed 6 December 2009

Beckenbach F, Briegel R, Daskalakis M (2009) Evolution and dynamics of networks in 'Regional Innovation Systems' (RIS). In: Pyka A, Scharnhorst A (eds) Innovation networks. Springer, Berlin

Chiaromonte F, Dosi G (1993) Heterogeneity, competition, and macroeconomic dynamics. Struct Chang Econ Dyn 4(1):39–63

Clarke LE, Weyant JP (2002) Modeling induced technical change: an overview. In: Grübler A, Nakicenovic N, Victor DG (eds) Technical change and the environment. Resources for the Future, Washington, pp 320–363

Cyert RM, March JG (1992) A behavioral theory of the firm. Blackwell, Malden

Daskalakis M, Kauffeld M (2007) On the dynamics of knowledge generation and trust building in regional innovation networks. A multi method approach. papers on agent based economics. http://www.ivwl.uni-kassel.de/beckenbach/poabe.html. Accessed 6 December 2009

Dawid H (2006) Agent-based models of innovation and technological change. In: Tesfatsion L, Judd KL (eds) Handbook of computational economics, vol 2. North-Holland, Amsterdam, pp 1235–1272

Dawid H, Reimann M (2011) Diversification: a road to inefficiency in product innovations? J Evol Econ 21(2):191–229

Dawid H, Reimann M, Bullnheimer B (2001) To innovate or not to innovate? IEEE Trans Evol Comput 5:471–481

Dosi G (1991) The research on innovation diffusion: an assessment. In: Nakicenovic M, Grübler A (eds) Diffusion of technologies and social behavior. Springer, Berlin, pp 179–208

Dosi G, Marsili L, Orsenigo L, Salvatore R (1995) Learning market selection and the evolution of industries. Small Bus Econ 7:411–436

Endres AM, Woods CR (2010) Schumpeter's 'conduct model of the dynamic entrepreneur': scope and distinctiveness. J Evol Econ 20:583–607

Fagiolo G, Dosi G (2003) Exploitation, exploration and innovation in a model of endogenous growth with locally interacting agents. Struct Chang Econ Dyn 14:237–273

Geroski PA (2000) Models of technology diffusion. Res Policy 29:603–625

Gerybadze A (1982) Innovation, Wettbewerb und Evolution: Eine mikro- und mesoökonomische Untersuchung des Anpassungsprozesses von Herstellern und Anwendern neuer Produzentengüter (Innovation, competition and evolution: a microeconomic and mesoeconomic investigation of adaptation processes of producers and users of new producer goods). JCB Mohr, Tübingen

Gort M, Klepper S (1982) Time paths in the diffusion of product innovations. Econ J 92 (September):630–653

Griliches Z (1957) Hybrid corn: an exploration in the economics of technological change. Econometrica 48:501–522

Grübler A, Nakicenovic N, Victor DG (1999) Dynamics of energy technologies and global change. Energy Policy 27:247–280

von Hippel E (1990) Task partitioning: an innovation process variable. Res Policy 19(6):407–418

Jaffe AB, Newell RG, Stavins RN (2003) Technological change and the environment. In: Mäler KG, Vincent JR (ed) Handbook of environmental economics. Elsevier Science BV, Amsterdam, pp 461–516

Kemp R (1997) Environmental policy and technical change. Edward Elgar, Cheltenham

Klimoski R, Karol BL (1976) The impact of trust on creative problem solving groups. J Appl Psychol 6:630–633
Levinthal D, March JG (1982) A model of adaptive organizational search. J Econ Behav Organ 2:307–333
Llerena P, Oltra V (2002) Diversity of innovative strategy as a source of technological performance. Struct Chang Econ Dyn 13:179–210
Lorenz E (1999) Trust, contract and economic cooperation. Camb J Econ 23:301–315
Mansfield E (1968) Industrial research and technological innovation. Norton, New York
March JG (1991a) How decisions happen in organisations. Hum-Comput Interact 6:95–117
March JG (1991b) Exploration and exploitation in organizational learning. Organ Sci 2(1):71–87
March JG (1994) A primer on decision making - how decisions happen. The Free Press, New York
March JG (1999) The pursuit of organizational intelligence. Blackwell, Malden
March JG (2006) Rationality, foolishness, and adaptive intelligence. Strateg Manage J 27 (3):201–214
March JG, Shapira Z (1987) Managerial perspectives on risk and risk taking. Manage Sci 33 (12):1404–1418
March JG, Simon H (1993) Organizations. Blackwell, Cambridge
Marengo L, Dosi G, Legrenzi P, Pasquali C (2000) The structure of problem-solving knowledge and the structure of organizations. Ind Corp Change 9:757–788
McEviliy B, Zaheer A (2006) Does trust still matter? Research of the role of trust in interorganizational exchange. In: Bachmann R, Zaheer A (eds) Handbook of trust research. Edward Elgar, Cheltenham, pp 280–302
Metcalfe S (1988) The diffusion of innovations: an interpretative survey. In: Dosi G (ed) Technical change and economic theory. Pinter Publishers, London
Morrison K, Potts J (2009) Toward behavioural innovation economics - heuristics and biases in choice under uncertainty. http://www.uq.edu.au/economics/abstract/379.pdf. Accessed 6 December 2009
Nelson RR, Winter SG (1982) An evolutionary theory of economic change. Harvard University Press, Cambridge
Noteboom B (1992) Towards a dynamic theory of transactions. J Evol Econ 2(5):281–299
Noteboom B (2004) Inter-firm collaboration, learning and networks: an integrated approach. Routledge, London
Noteboom B, Haverbeke van WPM, Duijsters GM, Gilsing VA, Oord vd A (2007) Optimal cognitive distance and absorptive capacity. Res Policy 36:1016–1034
Okada T, Simon HA (1997) Collaborative discovery in a scientific domain. Cogn Sci 21 (2):109–146
OECD, Eurostat (2005) Oslo manual: guidelines for collecting and interpreting innovation data. http://browse.oecdbookshop.org/oecd/pdfs/browseit/9205111E.PDF. Accessed 6 December 2009
Rammer C (2006) Gespaltene Innovationskonjunktur. ZEW-News, March. ftp://ftp.zew.de/pub/zew-docs/zn/Innovationen0306.pdf. Access 6 December 2009
Rogers EM (1995) Diffusion of innovations. The Free Press, New York
Schwarz N, Ernst A (2009) Agent-based modelling of the diffusion of environmental innovations - an empirical approach. Technol Forecast Soc 76(5):497–511
Schwoon M (2006) Simulating the adoption of fuel cells vehicles. J Evol Econ 16:435–472
Scitovsky T (1976) The joyless economy. Oxford University Press, New York
Shackle GLS (1961) Decision order and time in human affairs. Cambridge University Press, Cambridge
Silverberg G, Verspagen B (1994) Collective learning, innovation and growth in a boundedly rational, evolutionary world. J Evol Econ 4:207–226
Silverberg G, Verspagen B (1995) Evolutionary theorizing on economic growth. http://www.iiasa.ac.at/Publications/Documents/WP-95-078.ps. Accessed 6 December 2009

Silverberg G, Dosi G, Orsenigo L (1988) Innovation, diversity, and diffusion: a self-organisation model. Econ J 98: 1032–1054
Simon AH (1997) Administrative behavior. A study of decision - making processes in administrative organizations, 4th edn. Free Press, New York
Utterback JM (1996) Mastering the dynamics of innovation. Havard Business School Press, Boston
Vromen JJ (2001) The human agent in evolutionary economics. In: Laurent JN, Nightingale J (eds) Darwinism and evolutionary economics. Edward Elgar, Cheltenham, pp 184–208
Windrum P, Birchenhall CT (2005) Structural change in the presence of network externalities. J Evol Econ 1:123–148
Windrum P, Birchenhall CT (2009a) Consumer heterogeneity and the development of environmentally friendly products. Technol Forecast Soc Change 76:533–551
Windrum P, Birchenhall CT (2009b) Environmental impact quality and price: consumer trade-off and the development of environmentally friendly technologies. Technol Forecast Soc Change 76:552–556
Witt U (1987) Individualistische Grundlagen der Evolutorischen Ökonomik (Individualistic foundations of evolutionary economics). JCB Mohr, Tübingen
Witt U (2009) Propositions about novelty. J Econ Behav Organ 70:311–320
Wittmann T (2008) Agent-based models of energy investment decisions. Physica, Heidelberg

Creative Production in the Creative Industries

Michael Hutter

Abstract Branches of the economy that depend strongly on a constant flow of novelty, namely those of the cultural and the creative sector, have played a central role in reversing the trend towards homogenization which, according to Schumpeter, threatened to stifle the innovative process.

The particular conditions leading to the production and consumption of novelties are discussed. Most of these goods are information goods that generate affective sensations in their users. As the goods are offered and accepted by a public, valuations of their quality are communicated among amateurs and experts. The combination of surprising information and devices of valuation frames the production process in the creative industries. An economy's capacity for processes that appreciate and depreciate new contributions, and its capacity to keep the accumulated novelty capital available in material and virtual archives, is decisive for its sustained growth.

1 On Creative Industries

In *Theorie der wirtschaftlichen Entwicklung* (1912), Schumpeter had argued that the flow of new consumption goods, expanded through new methods of production, as well as new market and corporate forms, causes old merchandise markets to wither, while continuously creating new investment opportunities. Thirty years later, in *Capitalism, Socialism and Democracy* (1942), he saw the same basic process at work, but circumstances had changed drastically: class traditions had disappeared, legal forms, such as contract law, had lost their binding character. Therefore, the economic processes "congeal" into large industrial production

M. Hutter (✉)
Berlin Social Science Center, Berlin, Germany

Technische Universität Berlin, Berlin, Germany
e-mail: mhutter@wzb.eu

G. Buenstorf et al. (eds.), *The Two Sides of Innovation*, Economic Complexity and Evolution, DOI 10.1007/978-3-319-01496-8_8,

systems, supplemented by routinized innovation systems. In this environment, the investment opportunities diminish, and Marx's prediction of the falling rate of profit appears to enjoy a late confirmation.

Schumpeter had culled his empirical evidence from the research for *Business Cycles* (1938), which he had just completed before he wrote *Capitalism, Socialism and Democracy*. The post- war spread of large scale factories in all branches of the economy, from the agricultural industry to automobile manufacturing, appeared to confirm the thesis of the self-destructing decentralized economy.

Since the 1970s, however, tendencies that turned the trend toward the mechanization and socialization of production processes into its opposite have asserted themselves. Of central importance were innovations in the communication technologies—semiconductors, computers with rapidly increasing memory and computing capacities, digitalized transmission networks, and, finally, mobile telephones and other devices that make communication easier. The new products gained an increasing portion of total market share. In addition, their full productivity potential was realized in all the industries in which they became integrated as a part of the technical equipment for production, and of the information infrastructure within commercial organizations.

It seems as though the branches of the economy that depend most strongly on a constant flow of novelty, namely those of the cultural and the creative sectors, have played a central role in reversing the trend towards homogenization which threatened to stifle the innovative process. These industries profited in many ways from the productivity gains in the telecommunication technologies. At the end of the nineteenth century, all artistic and musical products, from oil paintings to symphonies, were still produced manually. Since then, mechanized storage and reproduction devices have come into use for texts and for audiovisual formats. The products of the film, television, radio and music industries are now distributed worldwide. Several technical generations of carrier media have been employed, each of them triggering a boost in sales. The interconnectedness of computers with continually increasing processing power has further increased the volume of content traffic. At the same time, value generated in the traditional branches, for example, in the market for art objects, has also increased.[1] Recent reports on the "culture and creative economy" in dozens of countries have established that this sector ranks among the largest in developed economies, and that growth rates are higher than average.[2] In consequence, the sector has received particular attention.

The existence of the "creative industries" as a separate sector is not a natural condition, but rather an institutional and performative achievement (Schlesinger 2009). The beginnings can be traced back to the efforts of the New Labour government in Great Britain at the end of the 1990s. In 1998, the (at that time newly renamed) *Department of Culture, Media and Sports* identified 13 industries as components of a sector that was ascribed a special relevance for economic

[1] Report artprice.com 2009.

[2] See Söndermann et al. 2009; Unctad 2008.

growth.[3] The sector encompasses the traditional cultural industries (arts, crafts, performing arts and design), industries of mechanized entertainment goods (music, publishing, film and television), industries with digitized products (video games, electronic publishing and software programs), fashion and advertising services. This attempt to construct a new economic sector gained in relevance when other countries decided to adapt their statistical reporting to the new sector. To account for a "creative industries" sector poses difficult methodical decisions: the value of publicly and privately financed enterprises is aggregated, and the sector size is systematically exaggerated through the inclusion of products that may have the form of texts or program codes, but do not contain novelty. Despite these weaknesses, the data have been used to demonstrate that a new economic sector has emerged that contributes roughly 5 % to GDP (Unctad 2008).

There are also attempts to extend the sector even further to cover all the commodities designed to provide experiences to the user, rather than being useful in attaining some ulterior good (Hutter 2011). This larger sector would include gambling, pornography and cultural tourism. Such an agglomeration has explanatory value in the context of Schumpeterian development: If the size of this sector were large enough, then creative destruction—which generates added value—could prevail against destructive destruction—which decreases GDP.

At the center of the productivity gains of the creative sector is the characteristic of newness or novelty. But the effect of novelty reaches beyond this sector. Novelties open previously unknown domains of possibility. These domains lead to "novelty-induced uncertainty" (Witt 2009). Furthermore, it has been argued that newness-generating processes do not only occur in the creative industries, but that they are paradigmatic for the evolutionary development of the entire economy (Cunningham and Potts 2008; Potts et al. 2008).

All this, taken together, makes it worthwhile to find out what it is that drives the demand for novelty, and thus the productive dimension of creativity.

2 Newness as Surprises

Newness emerges in agents' heads, more precisely, in their minds. Minds are embedded in organic nervous systems, capable of differentiating between past and future events and of relating them to each other. In the mind, a sensation that we call surprise can be triggered by a theater performance, or a video game or an advertising campaign. Surprise, literally "being taken over", connotes a state of being unprepared for an unexpected turn of events. The newness of the surprise is part of an affective sensation—something that fascinates and thus attracts the user. Fascinating surprises can be staged in a number of different ways. A particularly

[3] See DCMS 1998. The exact number of branches contained in the definition has shifted slightly over the years.

potent source is amazement. Virtuosity and originality evoke unique aesthetic experiences in artistic performances, be they visual, musical or literary.

Classical uncertainty—modelled as distance from an ideal condition of certainty—is but one member in the set of uncertainties purposefully created by events containing new information. Shackle (1949) spoke of potential surprise in order to find a measure for the contingency of future knowledge. Luhmann (1997) equated new information with surprising understanding. Dosi and Winter (2002) pointed to the importance of surprises in evolutionary events and to the difficulties in modelling them. The common concern of each treatment is that events cannot be accommodated through the notion of uncertainty.

The business strategies of the creative industries suggest that it is sufficient to achieve quite moderate degrees of surprise. Each enhancement of experience is perceived by the mind in its difference to what is already known to it. Thus, surprises are only comprehensible in their difference to something familiar. The novelties consumed are actually "familiar surprises".

The internal, mental experience of familiar, yet fascinating surprises might be a subject for neuro-economic research. To the social scientist, only the social expressions connected with such commodities are accessible. Performances of experience goods take place in public spaces—either in real, material places or in the imagined publics of all those who have read a novel or watched a film and now talk about them. The performances in public places have economic and social reasons. The economic reason lies in the fact that the consumption of products consisting of meanings is non-rivaling, as the perception and its interpretation are both performed in the minds of all the agents who submit themselves to the content being performed. Experience products, then, are primarily public goods, even though access barriers can keep supply scarce enough to generate positive prices. The social reason lies in the external effects of expressions of amazement, enthusiasm or rejection. The experience of each individual agent has greater value if the positive reactions of other agents can be experienced at the same time. The reactions confirm one's own experience. At the same time, they determine the criteria for the choice of future products. Thus, the public places and invisible scenes are embedded in networks of social contacts (Uzzi and Spiro 2005).

Public spheres and scenes are found around performances, material or virtual. They manifest themselves in simultaneous or subsequent reactions of the audience to the performances. The audience may be present, as in a concert, or absent, as in the reading of a book. The reactions express valuations (Karpik 2010). A simple reaction-valuation is the intensity of the applause with which amateurs express their direct, emotional appreciation. More complex and enduring are the valuations by experts, who have experienced a wider range of performances, and are skilled in ranking them along several scales of quality to arrive at a differentiated valuation of a new performance. Valuations record positive or negative deviations from expected achievements. The experts succeed in making entire product series comparable with other such series. Thus, they build a valuation pyramid with a few excellent works at the top, and many insignificant works at its base. Such pyramids are then destroyed and reconstructed by subsequent generations of experts.

While this social process goes on, a totally different form of valuation takes place through the purchase of works or of access rights to new performances and their spin-off products. Purchases realize exchange values based on individual choices, but the choices have been shaped by the valuations current in the social networks. In other words: aesthetic valuations that refer to excellence generate the value realized in economic transactions.[4]

Valuations are a form of attributing meaning. The creation and continual changing of meanings is a collective act that takes place between minds, which thus form the nodes, of a network. A very powerful example is language (Herrmann-Pillath 2009). Language is the oldest human carrier of meaning, but the use of signs began much earlier. Today, it invites the use of all senses. Language is comparatively rigid in its "ascription" of meanings, while artificial signs, independent of physical sound production, have much broader means of expression. This historical development was promoted by the products of the creative industries, from printed books to recorded works of dance, music and images, and on to the intricate codes of meaning of modern audiovisual media.

Within these networks of meaning emerge the valuations that attribute more or less "quality" to any new content that provides experiences. As new experience goods are offered and accepted by the public, valuations of the quality of the surprise events are communicated among amateurs and experts. Because of the valuations, a few of the new variations are repeated in further events, imitated or given new references, and all the others are soon forgotten. The selected products win the competition in the context of the valuation communication. Because we are dealing with surprises, the interest in the selected products usually diminishes within a few weeks or even years. Only a very small percentage of the products will be retained by being continually reinterpreted, valuated anew and then reproduced in different media, for example, in educational programs or in illustrations in magazines. Such canonized works or "cultural icons" are in turn formative for future valuations. In the long run, they even shape the infrastructure of the system of meaning, namely the minds and brains of the agents who are continually searching for new individual and social impulses.

The system of meaning develops in populations of content events or surprise products by means of the evolutionary algorithm of variation, selection and retention. In principle, the system evolves blindly—the maxim "nobody knows" comes from the movie industry, because acceptance and success cannot be predicted due to the contingency of future meanings (Caves 2000). The sum of populations is as large as the sum of events in which meanings are exchanged and valuations are constructed. It encompasses processes that create meaning in families, political systems, art scenes and companies, just to mention a few examples. The value chain of the creative industries only forms one, though especially remarkable, concentration in these events. Because of the density and intensity of the process within

[4] The question of the anchoring of valuations using the concepts of infinity (praise value) and zero (price value) is developed in Hutter (2010). See also: Hutter and Frey (2010).

this sector, it becomes clear that the traditional linear chain of value agglomeration is substituted by a circular process: in the "value spiral", new ideas and scripts are thought of, tried out, turned into products, distributed and used, but then the process continues as the products are appreciated and turned into new ideas and scripts by some of the users (Hutter 2006). Remarkable in this regard is an epistemological shift in the relationship between individual agents and valorizing network: the network is observed in the context of (sign-producing) agents who contribute to the ongoing value spiral, rather than observing the agents in the context of a valorizing network.

The value-generating processes in the creative industries are not only remarkable in themselves, but they also have paradigmatic character. The dual value of creative industry products can be found in products of other economic sectors as well. They are also charged with new meanings and ranked on valuation scales. Innovations seldom have a productivity-increasing effect that is evident for everyone. The advantages of any novelty have to be asserted in elaborate promotion campaigns. To use creative industries products as paradigmatic cases makes it easier to perceive the dimension of valuation in other economic sectors.

Thus we return to the tension between creation and destruction: new products and processes continue to emerge, even at an increasing rate. The novelties are familiar enough to be reproduced by other agents but nevertheless surprising enough to provide changing experiences. The novelties not only make up for the continually dwindling investment possibilities in previously successful products, but they even allow additional possibilities to emerge.

3 Outlook: New Research Dimensions

The focus of attention on the productive dimension of novelty in the creative industries contributes to the development of evolutionary economics because it motivates a shift in research interest from producers to consumers, away from industries and their technical and institutional-organizational characteristics and towards the exploration of user networks. The suppliers of experience goods stage their performances, but the users of the goods contribute their judgments. An evolutionary theory explains the ongoing processes of creative destruction in terms of the variations of authors' contributions, but also their selection by producers as well as users, and their retention (or restabilization) in the value constellations of educational canons, expert judgments and bestseller lists. In such a view, an economy's capacity for processes that appreciate and depreciate new contributions, and its capacity to keep the accumulated novelty capital available in material and virtual archives, is decisive for its sustained growth.

As research interest shifts from the cognitive, neuronal and physiological predispositions of humans towards their social dispositions, the findings of art history and the history of science, the humanities, cultural sciences and social sciences become more pertinent relative to those of the natural and life sciences. Cultural processes, with their own kind of autonomous self-reproduction in rituals and routines, appear to dominate the earlier biological processes. Patterns of preference are determined in communities of taste rather than in individual brains. In such communities, it is decided what will be considered a crazy idea, what will be a successful invention, and what will be a basic innovation.

References

Artprice (2009) Contemporary art market 2008/2009. The Artprice annual report. Artprice, London

Caves R (2000) Creative industries. Contracts between art and commerce. Harvard University Press, Cambridge, MA

Cunningham S, Potts J (2008) Four models of the creative industries. Centre of Excellence for Creative Industries and Innovation, Brisbane

Department of Media, Culture and Sports (1998) Creative industries mapping document. Department of Media, Culture and Sports, London

Dosi G, Winter S (2002) Interpreting economic change: evolution, structure and games. In: Augier M, March J (eds) The economics of choice, change and organization. Edward Elgar, Cheltenham, pp 337–353

Herrmann-Pillath C (2009) The economics of identity and creativity: a cultural science approach. University of Queensland Press, Brisbane

Hutter M (2006) Neue Medienökonomik. Fink, München

Hutter M (2010) Constant surprises. Value creation in the creative industries. In: Beckert J, Aspers P (eds) The worth of goods. Cambridge University Press, Cambridge, MA

Hutter M (2011) Experience goods. In: Towse R (ed) A handbook of cultural economics, 2nd edn. Edward Elgar, Cheltenham

Hutter M, Frey B (2010) On the influence of cultural value on economic value, in: Revue d'économie politique

Karpik L (2010) The economics of singularities. Princeton University Press, Princeton

Luhmann N (1997) Die Gesellschaft der Gesellschaft. Suhrkamp, Frankfurt

Potts J et al (2008) Social network markets: a new definition of the creative industries. J Cult Econ 32(3):167–185

Schlesinger P (2009) Creativity and the experts: new labour, think tanks and the policy process. Int J Press Polit 14(3):3–20

Shackle GLS (1949) Expectations in economics. Gibson, London

Söndermann M et al (2009) Gesamtwirtschaftliche Perspektiven der Kultur- und Kreativwirtschaft in Deutschland. Bundesministerium für Wirtschaft und Technologie, Berlin

UNCTAD (2008) Creative economy report 2008. United Nations, New York, NY

Uzzi B, Spiro J (2005) Collaboration and creativity: the small world problem. Am J Sociol 2005 (111):447–504

Witt U (2009) Propositions about novelty. J Econ Behav Organ 70:311–320

Labor Market Integration Policies and The Convergence of Regions: The Role of Skills and Technology Diffusion

Herbert Dawid, Simon Gemkow, Philipp Harting, and Michael Neugart

Abstract We study the role of different labor market integration policies on economic performance and convergence of two distinct regions in an agent-based model. Production is characterized by a complementarity between the quality of the capital stock and the specific skills of workers using the capital stock. Hence, productivity changes in a region are influenced both by the investment of local firms in high quality capital goods and by the evolution of the specific skill distribution of workers employed in the region. We show that various labor market integration policies yield, via differing regional worker flows, to distinct regional distributions of specific skills. Through this mechanism, relative regional prices are affected, determining the shares that the regions can capture from overall consumption good demand. There occurs a trade-off between aggregate output and convergence of regions with closed labor markets resulting in relatively high convergence but low output, and more integrated labor markets yielding higher output but lower convergence. Furthermore, results differ substantially in several respects as distinct labor market opening policies are applied.

Reprinted from Journal of Evolutionary Economics 22(3), 543–562, Springer (2012).
Electronic supplementary material The online version of this article (doi:10.1007/s00191-011-0245-1) contains supplementary material, which is available to authorized users.

H. Dawid (✉) • S. Gemkow • P. Harting
Department of Business Administration and Economics, Bielefeld University, Bielefeld, Germany
e-mail: hdawid@wiwi.uni-bielefeld.de

M. Neugart
Department of Law, Business, and Economics, Technical University of Darmstadt, Darmstadt, Germany

G. Buenstorf et al. (eds.), *The Two Sides of Innovation*, Economic Complexity and Evolution, DOI 10.1007/978-3-319-01496-8_9,

1 Introduction

An important challenge faced by the European Union (EU) is the integration of several Eastern European countries that became EU members at the eastward enlargement in 2004. Ten new countries joined the EU then, eight of which formerly were under communist rule. From the fall of the Iron Curtain to the present, these countries have faced similar structural problems that can be considered as an inheritance of the communist planned economy: a quantitatively and qualitatively worse endowment of physical capital, as well as less skilled human capital compared to established member states such as France, Germany and the UK.

In this paper, we ask (and try to answer), how different policies of opening up labor markets accompanying the integration process affect output and consumption in regions that start(ed) from different levels of economic development, and how these policies impact the convergence processes in general.

Prior to the eastward enlargement of the EU, a long-lasting and fierce debate emerged concerning the extent to which integration should also cover labor markets. The European Commission had to take into account in its decision the fear in the general public that an opening up of labor markets would lead through massive flows of labor from the east to the west to labor market tensions, social imbalances, and crime. In an influential policy paper, Boeri et al. (2002) argued that estimates of expected labor flows were showing a high degree of uncertainty, which would warrant a postponement of full liberalization of labor migration until more accurate numbers were available but by the same time would strongly suggest opening up labor markets, and phasing out the transition by a preannounced date. Finally, it was agreed that, after accession countries became part of the EU in May 2004, there should be a transition phase of up to seven years during which member countries would be allowed to impose restrictions on labor mobility. Along the "2+3+2" formula, restrictions had to be reviewed after two and then three years. Only the UK, Ireland and Sweden opened up their labor markets by 2004. After the first review, seven members lifted their restrictions, some simplified their procedures, while others, such as Austria and Germany, even after the second revision maintained their rigorous stance requiring permits for workers from former communist states.

Five years after the accession, it appears to be still to early to evaluate how economies of the old and new members states unfolded as a response to these particular labor market policy choices. Ultimately, it will be an empirical question to assess to what extent per capita growth on both sides of the former Iron Curtain, and convergence of the two so distinct European regions, were affected.

In fact, these policies, as they were implemented and in some cases changed accompanying the integration process of EU countries, raise important and so far not well understood research questions. In particular, it is an unresolved issue to what extent spatial frictions with respect to labor mobility may have positive or detrimental effects on overall and region specific variables related to the well-being of their citizens in the medium and long run. It is a non-trivial task to take into account the most likely consequences of various labor market integration policies in a world where there is complementarity between technologies used in a country and the specific skills needed to fully exploit the merits of these technologies, and where the speed of adoption of the necessary specific skills is driven by general skills, the distribution of which differs between countries.[1] Moreover, important feedback processes through wage dynamics, accompanying productivity growth and demand shifts driven by changes in households' consumption and the investment behavior of firms, interplay with the mobility of workers.

Our work relates to the growth literature on convergence, which, according to the neoclassical approach (see, e.g., Solow 1956; Mankiw et al. 1992; Barro and Sala-i-Martin 1992)), is the result of decreasing returns to physical or human capital. Among others, empirically the speed of convergence was investigated by Islam (1995) based on cross country data and by Canova and Marcet (1995) drawing on regional data. Howitt (2000) explains convergence taking cross-country knowledge spill-overs as a starting point, and Keller (2004) puts geographic proximity at the forefront.

We contribute to the convergence literature by studying the dynamic linkages between inter-regional spillovers through labor mobility and technological diffusion driven by demand induced investment of firms. All firms have access to latest technology provided by an investment-good producer. However, regions differ in their initial current productivity, which is a central distinguishing feature between new and old member states of the EU, or the West and former East Germany. In particular, there are different endowments with respect to the current technology used, the specific skills necessary to operate the capital stock, and the distribution of general skills. Although the latest technology can be bought by any firm in any region, there are restrictions on the use of this technology because of the constraining factor of specific skills. This brings into the picture a so far not well studied channel (see, e.g., Acemoglu 2009, Ch. 18) for the diffusion of technology and convergence of regions. Labor movement between the regions changes the allocation of specific and general human capital and thus the exploitability of the current technological level and the speed of adjustment to the current technological level. Choices of workers to commute are influenced by an exogenous and politically determined level of labor market integration. And thus, different policies on labor market integration may have non-trivial repercussions on productivity growth in the regions. Under particular consideration of the dynamic effects discussed above, we will address the following policy questions:

- How does the timing of the opening up of labor flows between (technological) leader and laggard regions affect growth and productivity development in both regions?

[1] There is ample empirical evidence that, in many cases, the diffusion of innovations requires adequate skills of the workforce of the firms adopting the innovation. (See, e.g. Bassanini and Scarpetta 2002; Griffith et al. 2004).

- What differences arise in short-term and long-term effects of different labor market opening policies?
- How far are technological spillovers induced by labor flows essential for the convergence between the regions?

There is little work on knowledge spillovers driven by migrating workers. Among the few suggestions elaborating that particular channel are Wong and Yip (1999) and a very recent study by Aghion et al. (2009). This investigates the spillovers occurring between regions in terms of human capital endowments by letting workers migrate. In that sense, our set-up is similar. We, too, look into the effects of different allocations of workers and therefore human capital across the regions by studying distinct scenarios of opening up regions for labor commuting. However, in terms of modeling choices, we look into a different role of human capital for the growth mechanics, as opposed to Aghion et al. (2009). In particular, they make a distinction between "high brow" and "low brow" educational endowments, with the former fostering innovation and the latter imitation of existing technologies. The role which we assign to human capital endowments is different. Our distinction is between general and specific skills, with general skills driving the speed of specific skill adoption which are necessary to run the current capital stock. Consequently, labor movements from one region to another have an effect on the specific and general skill levels in the regions, thus affecting the speed of adoption of the current level of productivity to the technological frontier.

Agent-based models have been developed in many areas of economics. They have been used to study the emergence of trading behavior on goods-markets and on financial markets, bidding behavior in auctions, numerous issues concerning innovation and industry evolution, and the emergence of cooperative behavior in economic systems. A large part of this work is surveyed in Tesfatsion and Judd (2006). Agent-based work in the area of macroeconomic modeling is, however, sparse. Closed macroeconomic models using an ACE approach have been provided, for example, by Chiaromonte and Dosi (1993), Silverberg and Verspagen (1993), Delli Gatti et al. (2005), Dosi et al. (2010) or Haber (2008) but these models focus neither on spatial aspects nor on the effects of labor flows. Closest to this work are the studies by Dawid et al. (2008, 2009), where the effects of different skill upgrading policies on technological change and growth were analyzed in the EURACE macroeconomic model, which is also the basis for this paper. The research focus of this paper on the comparison of labor market opening strategies, however, is quite distinct from the questions addressed in these previous studies. Also, in this paper, we deal with the interaction between regions which differ in several important aspects, whereas, in Dawid et al. (2008, 2009), scenarios were considered in which regions differ only with respect to the general skills distribution among workers.

On a more general level, the discussion of the policy experiments highlight the fact that an agent-based approach, to evaluate single or combined policy measures in the framework of a closed macroeconomic model with micro foundations that encompasses the interaction between different sectors, allows insights that go beyond the current economic literature. In particular, in the policy area considered here, we extend the literature reviewed above quite fundamentally by incorporating the feedback effects arising from technology and skills development through (regional) demand dynamics on consumption and investment goods markets, and by discussing the way frictions on different markets influence policy effects. Furthermore, we can explicitly distinguish between short and long run implications of policies.

Finally, our work is, in several respects, closely related to the literature in Evolutionary Economics. First, the driving factor of (regional) economic growth in our model is technological change brought about by a combination of investment by firms in new technologies and skill acquisition by workers. Second, our work is related to evolutionary industry life cycle models (e.g. Dosi et al. 1995) in a sense that the industry structure is endogenous and its evolution is driven by the relative competitiveness of firms caused by their investment in new technologies and by their ability to use these technologies. Third, our rule-based approach to capture firm behavior builds strongly on the rich evolutionary literature in this area such as, e.g. Nelson and Winter (1982) or Malerba et al. (2001).

We proceed by describing the core modeling assumptions, the parametrization of the agent-based model and the set-up of the experiments. In Section 4, we present and discuss our results from the various policy scenarios. In the last section, we conclude and sketch some directions for future analysis.

2 The Model

We conduct our analysis with an agent-based macroeconomic model that has a distinct regional dimension. Rather than fully describing in detail the various elaborate features of our framework, we stick here to a description of the core assumptions necessary for an understanding of our results. The model, already used for other policy analysis on the role of fostering human capital endowments in a spatial context, is explained in detail in the Online-Appendix A. (See also Dawid et al. 2008, 2009).

The focus in our experiments on labor market policy integration is on the interaction of three markets namely the labor, the consumption goods, and the capital goods market in a regional context, i.e. each firm and each household is located in one of the regions. The spatial extensions of the markets differ. The capital goods market is global meaning that firms in both regions buy from the same capital good producer and therefore have access to the same technology. On the consumption good market demand is determined locally in the sense that all consumers buy at a regional market located in their region, but supply is global because every firm might sell its products in all regional markets of the economy. Labor markets are characterized by spatial frictions determined by commuting costs that arise if workers accept jobs outside their own region. We model commuting costs as a proxy for the various degrees of labor market integration, where these costs may be inhibitive so that no worker flows occur, or may be at more moderate levels at which workers decide to accept jobs in the other region if the wage difference nets out the costs from commuting. The basic time unit in the model is one day, where many decisions, such as production choice or hiring by firms, are take monthly.

The consumption goods producer uses labor (L) and capital (K) as input factors. Both are vertically differentiated. The production quantity $Q_{i,t}$ of firm i in period t is given by

$$Q_{i,t} = \min\left[B_{i,t}, A_{i,t}\right] \times L_{i,t}^{\alpha} K_{i,t}^{\beta}, \tag{1}$$

where $B_{i,t}$ denotes the average specific skill level in the firms, $L_{i,t}$ is the number of workers, and $\alpha + \beta = 1$. The variable $A_{i,t}$ measures the average quality of the capital stock of firm i at time t. Note that, due to min $[B_{i,t},A_{i,t}]$, there is complementarity between the quality of capital goods and the specific skill level of the workers. The average quality of the capital stock of a firm increases over time; due to investments of the firm, the most recent vintages of the capital good are added to the stock. The technological quality of the capital good sold by the capital good producer increases over time following a random process, and the quality of the capital good sold at time t is referred to as the 'technological frontier' at time t.

Workers' human capital endowments have two dimensions. They embody an exogenously given level of general skills and an endogenously level of specific skills which changes on-the-job with the operation of the currently employed technology. The acquisition of specific skills in production is faster for higher general skill levels. The specific skills can be interpreted as capabilities and experiences obtained on the job. These skills are associated with the technology being used by the employer. Formally, the workers increase the specific skills over time by a learning process. The speed of learning depends on the general skill level b_w^{gen} of the worker w and the quality of the technology $A_{i,t}$ used by employer i.

$$b_{w,t+1} = b_{w,t} + \chi(b_w^{\text{gen}}) \max\left[0, \left(A_{i,t} - b_{w,t}\right)\right] \tag{2}$$

Here $b_{w,t}$ are the specific skills of worker w in period t and $\chi(b_w^{\text{gen}})$ increases with b_w^{gen}. The average specific skills $B_{i,t+1}$ in firm i is given by the values of $b_{w,t+1}$ averaged over all workers employed by firm i at $t + 1$.

The wage offer has two constituent parts. The first part is the market driven base wage $w_{i,t}^{\text{base}}$. The base wage is paid per unit of specific skill. If the firm cannot fill its vacancies, it increases the base wage to attract more workers. The second part is related to the specific skills. Since the specific skills represent the productivity of the workers, the wage $w_{i,t}$ is higher for higher specific skills. For each of the general-skill groups, the firm i offers different wages $w_{i,t,g}$ in period t. The wage offers are given by

$$w_{i,t,g} = w_{i,t}^{\text{base}} \times \bar{b}_{i,t,g} \tag{3}$$

where $\bar{b}_{i,t,g}$ are the average specific skills of all workers with general skill g in the firm. The underlying assumption of this determination of wage offers is that firms can observe general but not specific skills of job applicants.

The price of the capital good increases with its rising quality and, finally, the price of the consumption good sold by any of the firms in the market is determined by a standard elasticity based pricing rule. Assuming that all firms have constant expectations $\varepsilon_i^e < -1$ of the elasticity of their demand, they set the price according to the rule

$$p_{i,t} = \frac{\bar{c}_{i,t-1}}{1 + 1/\varepsilon_i^e}, \tag{4}$$

where $\bar{c}_{i,t-1}$ denotes unit costs in production of firm i in the previous period. As indicated above, producers distribute their goods by offering them at the posted price $p_{i,t}$ at the markets located in each region. The quantities they deliver to each of these markets every month (and hence the total production quantity in that month) are determined by production planning heuristics applied to demand estimations based on past data. Consumers regularly visit their regional market and make purchasing decisions described by a standard logit-choice model. For these and numerous additional decisions not mentioned here (detailed explanations are in the Online-Appendix A), the general modeling approach is to find rules which are backed by managerial decision rules documented in the corresponding management literature, or empirically based results on consumer choices that can be found in the corresponding marketing literature. Overall, the modeling choices are when feasible, empirically based, as is the parametrization to which we turn now.

3 Parametrization and Set-up of Experiment

Table 1 summarizes the general set up in terms of the numbers and types of agents and regions involved. There are two regions: each region hosts 800 households, 40 consumption good producers, and a regional market denoted as mall. There is a single capital goods producer.

We model the two regions featuring distinct total factor productivity. It is not our intention to focus on two particular regions when choosing parameters and initializations of the simulations, as we want to make a more general point on the role of labor market integration policies on economic performance and convergence of two regions with different endowments of physical and human capital. However, to have a solid empirical grounding of the experiment, we choose values reflecting differences between Germany and Poland that are particularly relevant for the aspects of the integration dynamics we want to study. To that end, we rely on institutional and OECD data as well as on empirical findings reported in Growiec (2008). In this paper the distance to the world technological frontier is determined for OECD and new EU

Table 1 General set up

Description	Value
Regions	2
Households	1,600
Consumption goods producers	80
Capital goods producers	1
Malls	2

member countries based on Data Envelopment Analysis. Furthermore, Growiec (2008) uses the resulting estimates of relative efficiency levels to decompose the ratios of per capita output into ratios of several factors including physical capital per capita and human capital. We approximate the estimated ratios of physical capital stocks and of total factor productivity between Germany and Poland reported in Growiec (2008) by initializing the per capita stock of physical capital in the high income region 1 as three times higher than that in region 2 and both specific skill levels and average capital quality in region 1 at 150% of that in region 2. Furthermore, it is assumed that at, $t = 0$, the technology used in the region 1 corresponds to the technological frontier, and wages in region 1 are 1.8 times higher than in region 2 (Table 2).

In addition to the differences in the initialization of key variables in the two regions, we also capture institutional, respectively more persistent regional differences, by setting (constant) parameters differently in the two regions. In particular, we incorporate differences in general skills of workers between the regions as measured in the International Adult Literacy Survey (IALS) and represent differences between the social security systems by setting the wage replacement rate in case of unemployment to 70% in region 1 and 60% in region 2.[2] All other parameters were chosen as in previous calibrations of the corresponding simulation model (see Dawid et al. 2008, 2009) combining what the relevant literature reports on empirical estimates related to the various parameters of the model, with considerations of viability of the model and its ability to reproduce standard stylized facts.

After choosing the parameters for the model, we compare in our experiments four policies, which read as follows:

"closed": Workers can only work in their domestic region.
"closed-1000-open-c": Workers can work in both regions after the first 1000 periods (50 months) and have to bear commuting costs.
"open-c": Workers can work in both regions, but face commuting costs.
"open": Workers can work in both regions, but do not face commuting costs.

These experiments are thought to address the policy question formulated in the Introduction.

For each scenario, we perform 38 single runs. Each single run represents 6,000 periods (days), which corresponds to 300 months, since we assume that each month has 20 (working) days. We allow for a transient phase of 2,000 periods (100 months)

[2] Again, these numbers were motivated by empirical observations in Germany and Poland.

Table 2 Experiment design with initial values for different variables for the high income region 1 and low income region 2

	Region 1	Region 2
Technological frontier	1.5	1.5
Per capita capital stock	3600	1200
Productivity capital stock	1.5	1.0
Specific skill level	1.5	1.0
Wage	1.8	1.0

before policies are applied in order to let the economy develop and to avoid starting effects. The transient phase is not part of the economic analysis in the following and, consequently, not shown in the figures. During the transient phase, technological progress is switched off, no worker flows occur and consumption goods are only delivered to the regional malls where production took place. After 2,000 periods, consumption and capital good markets are opened between the regions and results are shown from 60 periods later onwards, to let firms adjust their delivery volumes of consumption goods to the other region after the regional goods markets open.

Before we discuss the findings of our simulation analysis, it is useful explicitly to point out why a simulation approach is needed to study the effects in which we are interested. A meaningful examination of the dynamics of the agent-based model per se, which in a mathematical sense can be interpreted as a Markov process in a high-dimensional state-space, by analytical means is prevented by the complexity of the involved transition functions. So, any analytical treatment would have to rely on a mean-field approach, where only the dynamics of the first few moments of the distribution of variables, such as the specific skill level of workers, is captured. Given our approach to describe individual decision process by means of empricially founded and partly quite complex rule-based heuristics, a closed form formulation of such a mean-field model would already be quite challenging. But even if such a model could be formulated, it is quite obvious that important aspects of our model, which will be crucial for the mechanisms driving our results, could not be captured. Consider, for example, the interplay of the movement of workers between regions and the dynamics of skill acquisition. The decisions of workers to accept a job in a different region depends on their reservation wage, which again depends on their (general) skills and their job history. These two aspects influence the specific skill levels of the considered workers, which means that the micro-structure of the interaction leads to a systematic bias in the specific skill levels of commuting workers relative to the average specific skills in that region. Furthermore, the acquistion of specific skills by workers commuting from an ex-ante low-tech to an ex-ante high-tech region will be systemtically different from the average speed of skill acquisition in that region due to their larger skill gap. Since these workers are more likely to move back to firms in the low-tech regions once these firms have closed the wage gap, another systematic bias with respect to skill transfer between the regions occurs when these workers move back to firms in their home region. To capture such effects, an explicit representation of the micro-interactions is needed and, hence, mean-field models or even more so representative agent models, which are analytically tractable, are not feasible alternatives.

4 Simulation Results

We are interested in how different policies targeting the integration of the two distinct regions fare. As a measure of performance, we consider (regional) output of the consumption good for most of our analysis. However, at the end of the section, we will also comment on the effects the different policies have on regional consumption. Figure 1 displays output using box plots that represent the distribution across 38 batch-runs for each policy scenario, where output is given by the average over the last twenty monthly observations. The four scenarios are ordered in a way that, from left to right, the amount of spatial labor market frictions go down. From the aggregate point of view, we observe that opening labor markets increases total output, but the particular way the labor market is opened has little long-run impact. Total output is lowest for the policy option of never allowing labor flows between the regions ("closed"). For the other three scenarios, which open up the labor market delayed by 50 months and impose commuting costs ("closed-open-1000-c"), open up immediately imposing some commuting costs ("open-c"), and full integration right away ("open"), no output differences occur. Applying the Wilcoxon rank test the only statistically significant differences occur if we compare the closed scenario with any of the integration scenarios.

Interestingly, the aggregated point of view hides regional differences occurring along the policy scenarios, again taking output as the performance measure. Figure 2 summarizes the outcomes, now showing box plots by policy scenario and by region. While output leveled off at the aggregate, it shows quite striking regional differences along the different policy options. For the high income region, as we start allowing for regional labor market flows going from the "closed" scenario to the "closed-open-1000-c" and the "open-c" scenario, output increases, again measured as the mean of the last 20 months of the simulated time series. Going for the full integration policy with no commuting costs involved yields an outcome located somewhere between the output levels of a closed and delayed opening of the regional labor markets.

An inverse ranking along the policy experiments can be observed for the low income region 2. For the first three policy scenarios output, becomes lower and lower as we integrate regional labor markets. Going to the extreme of imposing no commuting costs we get, again, an output level somewhere located between the output associated with a closed and regional labor markets opened after 50 months. Thus, looking at convergence of the two regions as a function of the various policy options, we are confronted with the least unequal distribution of output across regions if labor markets are closed, where, however, recalling the result from Fig. 1, total output was lowest, too. As among all the remaining polices, total output is equal, a ranking of the other three policies based on an objective to reduce

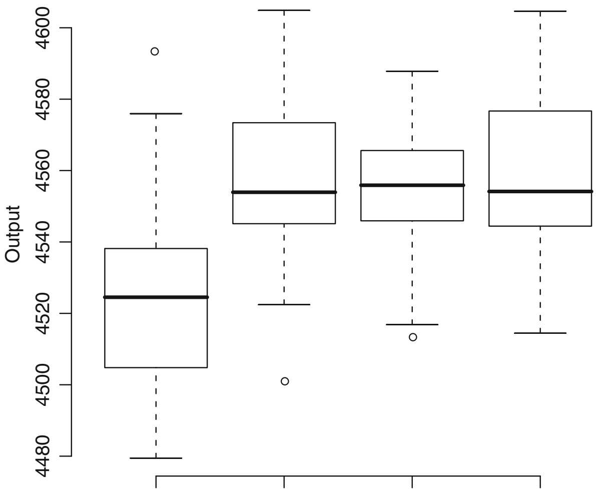


Fig. 1 Output (from *left* to *right*) for scenarios: “closed”, “closed-open-1000-c”, “open-c”, and “open”

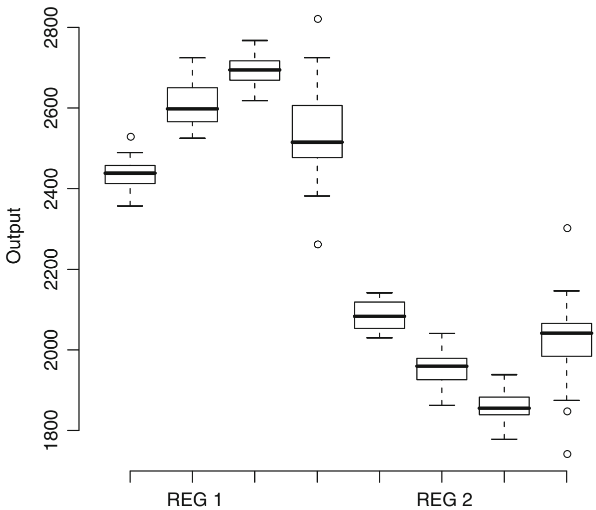


Fig. 2 Output by regions (from *left* to *right* within high income region 1 and low income region 2, respectively) for scenarios: “closed”, “closed-open-1000-c”, “open-c”, and “open”

regional inequality, would be “open” followed by “closed-open-1000-c”, and “open-c”. Two sided Wilcoxon signed rank tests were carried out for each pairwise comparison of policies in each region, and it was established that all observed output differences were statistically significant at a 99% level.

In the Online-Appendix B, we show that the qualitative features of Figs. 1 and 2 stay intact for variations of key parameters, such as the general skill levels in the two regions, the initial quality of the capital stock in the two regions and the location of the technological frontier. The crucial point is that the two regions differ with respect to the initial distributions of specific skill levels of workers and the initial quality of the capital stock such that there is a low income and a high income region. Accordingly, the mechanisms discussed below are relevant not only for the particular parametrization discussed here, but for a larger set of scenarios.

Figure 2 illustrates the long run effects of the different policies in the different regions, but, as can be seen in Fig. 3, where the dynamics of output produced in each region (averaged over the batch runs carried out for each scenario) are depicted, short run effects do not fully coincide with these observations. In particular, for the scenario where labor markets stay closed, the short run effects differ significantly from the long run effects. Short-run output in region 1 under closed labor markets is relatively high compared to the scenarios with labor market opening, whereas in region 2, for the first 100 months, output in the closed scenario is below that of the three scenarios with labor flows. The relative advantage of a closed labor market compared to an open one for region 2 emerges more than 100 months after the introduction of the different policies.

The remainder of the section will trace the mechanisms causing the regionally dispersive effects of the various labor market integration policies. Using the possibilities offered by micro-founded agent-based modeling, we will thereby illustrate that the effects of the policies are determined by particular feedbacks between price driven demand effects and flows of workers, know-how and capital. Essentially, what we are going to show by looking into the evolution of various region specific variables is that, due to cost and price differences between producers, demand in both regions shifts toward goods produced in one of the regions. This induces an increased demand for labor in that region, which under closed labor markets results in increasing wages, thereby reducing the local cost advantage. At the same time, investment in that region goes up, at least in the short run. Under open labor market scenarios, the increased labor demand in the region with initial cost advantages leads to labor flows, which on the one hand, induce technological spillovers between the regions, and on the other hand, alter the tightness of the two regional labor markets and imply quite different wage dynamics compared to the closed scenario. These countervailing effects drive region specific production costs and, ultimately, the relative prices, which in turn determine future worker flows.

Figure 4 shows the relative prices relating the price level of goods in region 2 to the price level of goods in region 1. Focusing on the last months of the time series, one sees how the convergence results coincide with the relative prices. As Fig. 4 reveals the "closed scenario", which is related by the solid line, has the lowest relative price levels by the end of the simulation period. The other relative price levels rank across the policy scenarios in accordance with the output levels across the two regions (see Fig. 2), with the "open-c" policy featuring the highest relative price and, consequently, the largest output gap between the two regions. Furthermore, it can be clearly seen that the difference between short- and long-run output effects of the closed policy is based on the fact that, in the short run, relative prices of goods produced in region 2 compared to those from region 1 go up before they start declining.

Digging deeper, trying to understand what is driving the relative regional price levels, one has to recall, first, that firms set prices as a mark-up on the average costs, with labor costs being a large share. Figure 5 illustrates the role of labor costs per

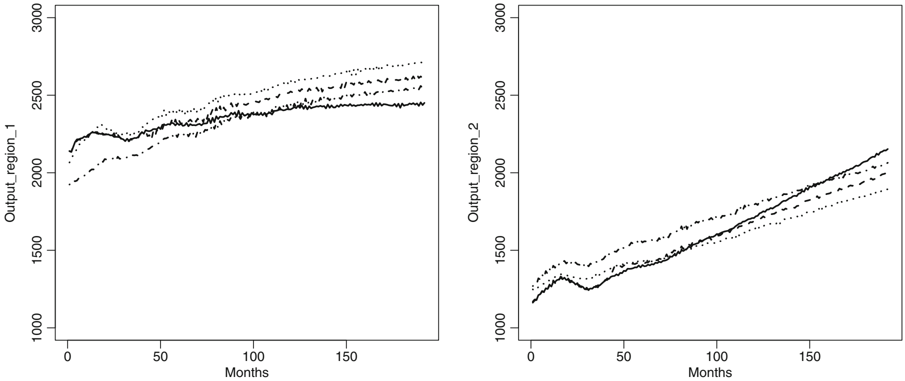


Fig. 3 Dynamics of output in region 1 (*left panel*) and region 2 (*right panel*) for scenarios: "closed" (*solid line*), "closed-open-1000-c" (*dashed line*), "open-c" (*dotted line*), and "open" (*dashed-dotted line*)

output by showing the relative wages between the two regions for the various policy scenarios and the relative specific skills, respectively. Specific skills constrain the firm using the available technology, and thus higher specific skills allow for higher production by making use of the qualitatively better capital stock. A first observation based on these two figures is that, in the open scenario, where the two labor markets are completely integrated without commuting costs at the same time when goods markets open up, the convergence between the two regions with respect to productivity and wages of workers employed in each region is almost perfect after 200 months. If labor flows are inhibited by spatial frictions, the degree of convergence is substantially reduced but still larger than in the case of closed labor markets. The "closed" policy yields the lowest relative wages measured as the wage level in region 2 to the wage level in region 1. Although one finds the lowest relative skill level for that same policy scenario, comparing the two figures makes clear that, in terms of relative specific skills, region 2 under the closed scenario comes close to the level it would reach under the open scenario, whereas the gap with respect to wages between these two policy scenarios remains much larger. The reason for this difference is highlighted in Fig. 6, where the "base wage" offer, i.e. the wage a firm pays per expected unit of specific skills of a worker, in the region 2 relative to that in region 1 is shown. Initially base wages in region 1 are larger, due to tighter labor markets in that region prior to the opening of goods markets, but in all three scenarios, where labor markets are opened, the relative base wages in region 2 go up over time and, for "open" and "open-c" policies, base wage offers in region 2 eventually exceed those in region 1. Only in the case of a closed labor market do the base wage offers in region 2 consistently decrease compared to those in region 1. These observations show that an important explanatory factor of the differences in relative prices under the four scenarios is the development of relative base wage offers. Comparing Figs. 4 and 6 shows, however, that other effects must be relevant, since, for a large time interval, relative prices in the open

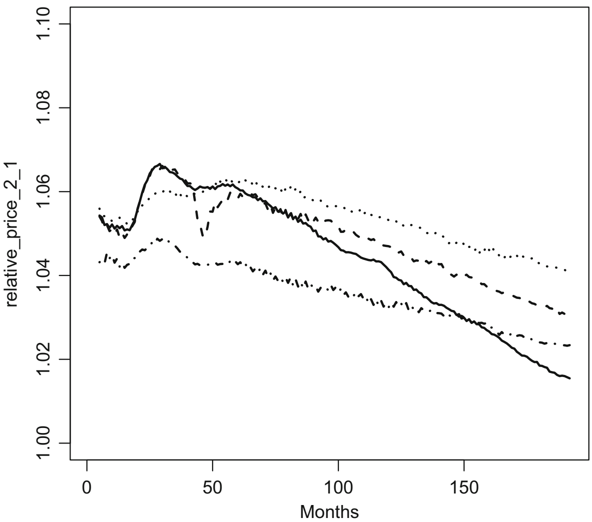


Fig. 4 Relative prices region 2 to region 1 for scenarios: "closed" (*solid line*), "closed-open-1000-c" (*dashed line*), "open-c" (*dotted line*), and "open" (*dashed-dotted line*)

scenario are below those in the closed scenario, although relative base wage offers in the open scenario always stay above the relative base wage offers under the closed policy. Before we return to that issue, we investigate further the reasons for the diverging dynamics of relative base wage offers in the four scenarios.

Obviously, both the dynamics of the base wage offers as well as the distributions of the specific skills by region and differentiated along the policy scenarios are strongly linked to the flows of workers between the regions. The two panels in Fig. 7 illustrate this feature. Again, the solid line refers to the policy scenario of closed regional labor markets where no commuting takes place. As one chooses the policy option of opening up labor markets after a transition period and still imposing commuting costs afterwards ("closed-open-1000-c"), one generates an immediate increase in commuters from low income region 2 to the high income region 1 (see dashed line). As wages paid in region 1 are higher than wages in region 2, it pays off for the workers in region 2 to accept job offers from firms in region 1, even taking into account costs from commuting. Worker flows from region 1 to region 2 only slowly increase as the wage levels of the two regions get closer. The only striking qualitative difference between the "closed-open-1000-c" scenario and opening up labor markets immediately but imposing commuting costs ("open-c") is the immediate onset of commuter flows from region 2 to region 1, as depicted by the dotted line in the right panel of Fig. 7. Otherwise, worker flows between the regions behave similarly in these two policy scenarios. Combining the flows in both directions to compute net worker flows between the regions, it can be easily seen that, in all three open scenarios, there are net worker flows from region 2 to region 1, and that these net flows are largest under the "open-c" policy. It is now easy to understand that the worker flows are the driving force underneath the differences in base wage offer dynamics between the four policy scenarios. In the "closed" scenario, the increased demand for labor in region 1 triggered by the additional demand faced by region 1 producers after the opening of the labor

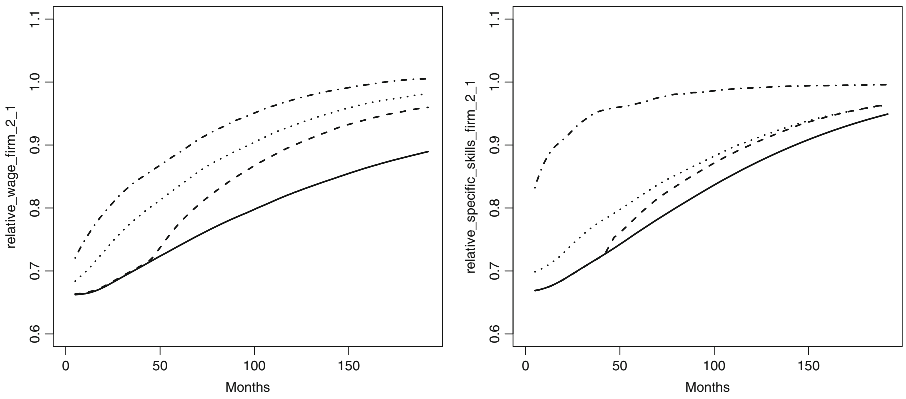


Fig. 5 Relative wages (*left panel*) and relative specific skills (*right panel*) of region 2 compared to region 1 for scenarios: "closed" (*solid line*), "closed-open-1000-c" (*dashed line*), "open-c" (*dotted line*), and "open" (*dashed-dotted line*)

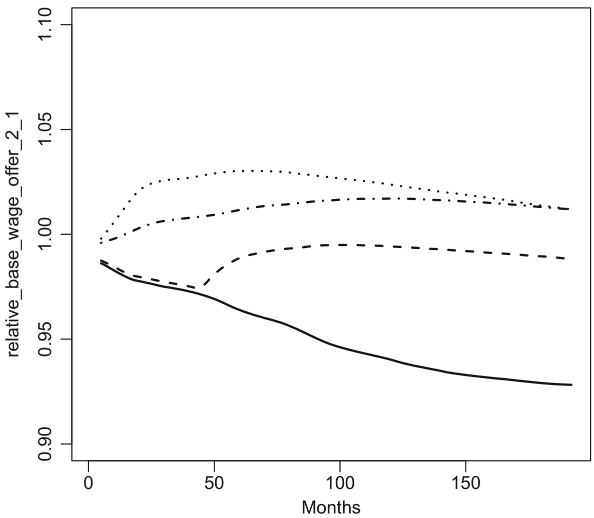


Fig. 6 Ratio of base wage offers in region 2 and region 1 for scenarios: "closed" (*solid line*), "closed-open-1000-c" (*dashed line*), "open-c" (*dotted line*), and "open" (*dashed-dotted line*)

market cannot be met by workers commuting from region 2. Therefore, producers in region 1 often have to deal with unfilled vacancies and, accordingly, increase their base wage offers more frequently than producers in region 2. Quite the contrary holds true for the "open" scenario, where producers in region 1 can easily attract workers from region 2. The resulting outflow of workers from region 2 increases the labor market tightness in that region, thereby generating an upward trend of base wage offers in that region.

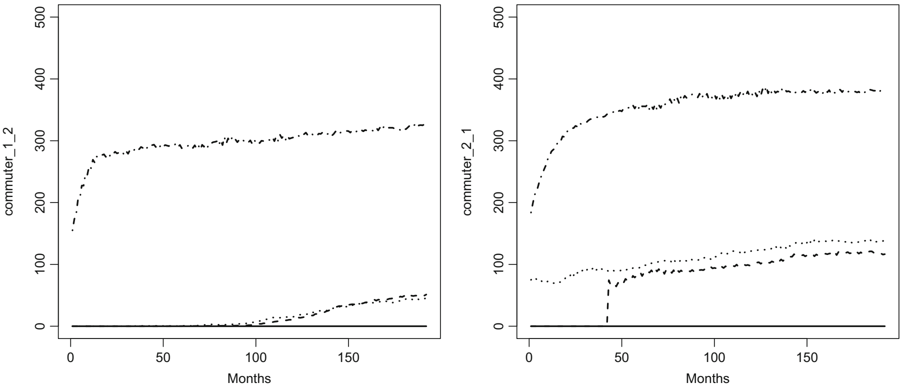


Fig. 7 Commuters from region 1 to region 2 (*left panel*) and commuters from region 2 to region 1 (*right panel*) for scenarios: "closed" (*solid line*), "closed-open-1000-c" (*dashed line*), "open-c" (*dotted line*), and "open" (*dashed-dotted line*)

As discussed above, in terms of the regional distribution of specific skills, the low income region 2 profits most from opening up the labor markets completely. Given the commuting patters the explanation for this observation is straight forward. In the open scenario, a large number of workers from region 1, who on average have higher specific skills than those in region 2, commute to region 2, whereas an even larger number of region 2 workers with relatively low specific skills commute to region 1, thereby lowering the average specific skill level of firms in there. Similar considerations explain the dynamics of specific skills in the other policy scenarios. An additional implication of the worker flows depicted in Fig. 7 is that the labor intensity of production in the two regions is affected by the applied policy, which, in turn, influences production costs and prices. Focusing again on region 2, due to the worker flows, the capital intensity of production under the "open" policy is larger than under the "closed" policy.[3] This implies that, per worker with a given specific skill level, more output is produced in the open than in the closed scenario, which explains our previous observation that relative prices in region 2 are lower in the "open" than in the "closed" scenario in many periods, although the relative base wage offers are always lower in the "closed" scenario.

Thus, the commuter flows induced by the different policies have intricate implications for specific skill distributions, wages and capital intensity, which influences prices, and thereby generates demand shifts which feed back on commuter flows and investments. As the relative prices finally determine the market share a region can ultimately capture, they explain the output and convergence patterns over the policies we analyzed.

[3] This observation has been checked considering the dynamics of capital/output ratios, but we refrain from presenting the corresponding figure here.

To finish our analysis, we wish to point out that the ranking of the different policies from the perspectives of the two regions was carried out with a focus on output produced by all firms located in that regions. A different perspective is to focus on the dynamics of the consumption of all households in a particular region. In scenarios with open labor markets, where many workers are employed outside their home region but still consume at home, the two perspectives might differ. Indeed, Fig. 8, which shows box-plots of consumption in both regions in the final 20 months, demonstrates that, for both regions, the policies that yield largest local production do not lead to highest local consumption. From the perspective of consumption in region 1, the best option is to keep labor markets closed, whereas, for region 2, the three scenarios in which labor markets are opened give higher consumption than the "closed" policy, although the ranking was exactly the opposite with resepct to local production. The reason for these differences is the interplay of commuter flows and relative wages in the two regions. Under the "open" policy, the large number of region 2 workers who commute to region 1 earn higher (real) wages than they would earn in their own region under the "closed" policy. This increases the overall consumption budgets of households in region 2 and, therefore, has positive effects on total consumption there. On the other hand, workers in region 1 profit from the larger real wages they earn in the "closed" scenarios compared to the three policies with labor market opening. Hence, total consumption in that region is largest in the "closed" scenario.

5 Policy Implications and Conclusions

The policy experiments on labor market integration yield strikingly different outcomes depending on the variable of interest, the regional level of analysis, and the time horizon. Therefore, policy implications on what labor market policies to choose will differ with the various objectives that may impose. We show that, while overall output is lowest for the closed scenario and equally higher for all policies that open up labor markets, output differs along all four policies if we look into the regional effects. Thus, while the policy advice is, when the objective is to maximize overall output, to choose either of the policies that at least gradually opens up labor markets, the advice to a policymaker who cares about convergence of regions would go differently. A policymaker who is willing to trade-off some output on the aggregate for more convergence should not integrate labor markets. If, however, a policymaker is not willing to give up overall output, then the advice is to open fully labor markets as among all the policies that promote labor market integration, as this is the policy which results in the least inequality between regions with respect to output levels (c.f. Figs. 1 and 2). In a world in which there are considerable flows of workers who work abroad but still consume in their domestic region, results in terms of convergence effects of the various labor market integration policies differ, and so will policy advice if a policymaker's objective is to reduce between region inequality regarding per-capita consumption (c.f. Fig. 8). With consumption as an argument in

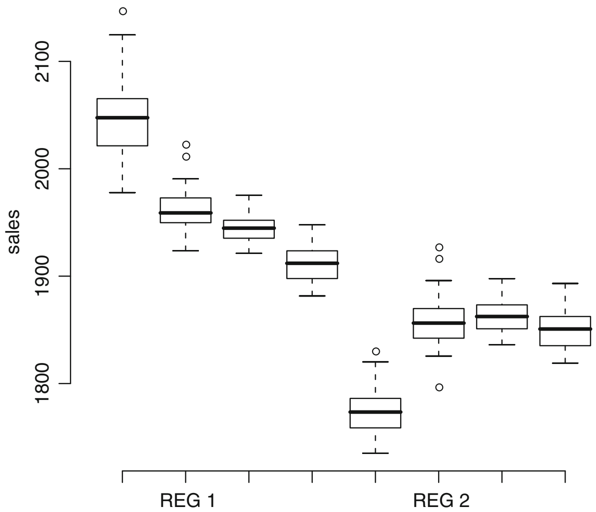


Fig. 8 Total consumption by regions (from *left* to *right* within region 1 and 2, respectively) for scenarios: "closed", "closed-open-1000-c", "open-c", and "open"

a policymaker's objective function, no trade-off between convergence of the regions and overall performance arises. Consequently, it is advisable to open up labor markets, as this policy yields better results in terms of overall consumption and convergence of regions than does one not allowing workers to commute between regions. Finishing up on the policy implications of our analysis, we want to remark that the implications drawn so far, differentiating along output or consumption, and the overall effects of policies as well as the effects on convergence, are implied by long-run outcomes. However, as illustrated in Fig. 3, the evolution of output differs with respect to the various policies by region. In particular, the high productivity region fares better initially in terms of output with a policy of closed labor markets. Thus, contrary to advice which relates to the long-run consequences, for a policymaker in that region who discounts the future heavily, it would be advisable not to open up labor markets.

Although the current model is already quite comprehensive, potentially interesting features are left out. Endogenous technological change would provide a richer model of productivity progress than our approach of using the standard assumption of an exogenously given process that eventually increases the quality of the capital good. A second feature that we will expand on in the future is the commuting behavior of the workers. So far, it is a very stylized approach based on a comparison of labor income net of some exogenously given costs for commuting. It would be a rewarding extension to put more structure on this particular decision of the workers, extending the framework to migration where workers actually reside in the region in which they chose to accept work. The framework we are using is sufficiently flexible to allow for these extensions, but they define a worthwhile agenda for future work.

Acknowledgements We are gratefully to an anonymous referee and the editors of this special issues for helpful comments. This work was carried out in conjunction with the EURACE project (EU IST FP6 STREP grant 035086) which is a consortium lead by S. Cincotti (University of Genova), H. Dawid (University of Bielefeld), C. Deissenberg (Université de la Mediterrané), K. Erkan (TUBITAK National Research Institute of Electronics and Cryptology), M. Gallegati (Università Politecnica delle Marche), M. Holcombe (University of Sheffield), M. Marchesi (Università di Cagliari), C. Greenough (STFC - Rutherford Appleton Laboratory).

References

Acemoglu D (2009) Introduction to modern economic growth. Princeton University Press

Aghion P, Boustan L, Hoxby C, Vandenbussche J (2009) The causal impact of education on economic growth: evidence from U.S. Harvard University, Mimeo

Barro RJ, Sala-i-Martin X (1992) Convergence. J Polit Econ 100:223–251

Bassanini A, Scarpetta S (2002) Does human capital matter for growth in OECD countries? A pooled mean group approach. Econ Lett 74:399–405

Boeri T, Bertola G, Brücker H, Coricelli F, Dolado J, Fitzgerald J, de la Fuente A, Garibaldi P, Hanson G, Jimeno J, Portes R, Saint-Paul G, Spilimbergo A (2002) Who's afraid of the big enlargement? Economic and social implications of the European Union's prospective eastern expansion, CEPR policy paper no. 7, CEPR London

Canova F, Marcet A (1995) The poor stay poor: non-convergence across countries and regions CEPR discussion paper no. 1265

Chiaromonte F, Dosi G (1993) Heterogeneity, competition and macroeconomic dynamics. Struct Chang Econ Dyn 4:39–63

Dawid H, Gemkow S, Harting P, Neugart M (2009) On the effects of skill upgrading in the presence of spatial labor market frictions: an agent-based analysis of spatial policy design. JASSS 12:4

Dawid H, Gemkow S, Harting P, Neugart M, Wersching K (2008) Skills, innovation and growth: an agent-based policy analysis. Jahrb Natlökon Stat/J Econ Stat 228:251–275

Delli Gatti D, Guilmi C, Gaffeo E, Giulioni G, Gallegati M, Palestrini A (2005) A new approach to business fluctuations: heterogeneous interacting agents, scaling laws and financial fragility. J Econ Behav Organ 56:489–512

Dosi G, Fagiolo G, Roventini A (2010) Schumpeter meeting Keynes: a policy-friendly model of endogenous growth and business cycle. J Econ Dyn Control 34:1748–1767

Dosi G, Marsili O, Orsenigo L, Salvatore R (1995) Learning, market selection and the evolution of industrial structures. Small Bus Econ 7:411–436

Griffith R, Redding S, van Reenen J (2004) Mapping the two faces of R&D: productivity growth in a panel of OECD industries. Rev Econ Stat 86:883–895

Growiec J (2008) Productivity differences across OECD countries. In: The World technology frontier revisited IBS working paper No.1/2008 institute for structural research, pp 1970–2000

Haber G (2008) Monetary and fiscal policy analysis with an agent-based macroeconomic model. Jahrb Natlökon Stat/J Econ Stat 228:276–295

Howitt P (2000) Endogenous growth and cross-country differences. Am Econ Rev 90:829–846

Islam N (1995) Growth empirics: a panel data approach. Q J Econ 110:1127–1170

Keller W (2004) International technology diffusion. J Econ Lit 42:752–782

Malerba F, Nelson R, Orsenigo L, Winter S (2001) Competition and industrial policies in a 'History-Friendly' model of the evolution of the computer industry. J Ind Organ 19:635–664

Mankiw GN, Romer P, Weil DN (1992) A contribution to the empirics of economic growth. Q J Econ 107:407–437

Nelson RR, Winter SG (1982) An evolutionary theory of economic change. Harvard University Press, Cambridge

Silverberg G, Verspagen B (1993) Collective learning, innovation and growth in a boundedly rational, evolutionary world. J Evol Econ 4:207–226

Solow R (1956) A contribution to the theory of economic growth. Q J Econ 70:65–94

Tesfatsion L, Judd KE (2006) Handbook of computational economics II: agent-based computational economics. North-Holland

Wong K, Yip KC (1999) Education, economic growth, and brain drain. J Econ Dyn Control 23: 699–726

Creative Destruction and Fiscal Institutions: A Long-Run Case Study of Three Regions

Lars P. Feld, Jan Schnellenbach, and Thushyanthan Baskaran

Abstract We analyze the rise and decline of the steel and mining industries in the regions of *Saarland*, *Lorraine* and *Luxembourg*. Our main focus is on the period of structural decline in these industries after the second world war. Differences in the institutional framework of these regions are exploited to analyze the way in which the broader fiscal constitution sets incentives for governments either to obstruct or to encourage structural change in the private sector. Our main result is that fiscal autonomy of a region subjected to structural change in its private sector is associated with a relatively faster decline of employment in the sectors affected. Contrary to the political lore, fiscal transfers appear not to be used to speed up the destruction of old sectors, but rather to stabilize them.

1 Introduction

Processes of creative destruction do not occur in a political vacuum. There is a vast literature on systems of innovation (see e.g. Freeman 1995 and Metcalfe 1995) which has taken into account the fact that innovative activities are influenced by policies and institutions that guide the interactions between those who participate in

Reprinted from Journal of Evolutionary Economics 22(3), 563–583, Springer (2012).

L.P. Feld
Walter Eucken Institut, Goethestr. 10, 79100 Freiburg i.Br., Germany
e-mail: feld@eucken.de

J. Schnellenbach (✉)
Ruprecht-Karls-Universität Heidelberg, AWI, Bergheimer Str. 58, 69115 Heidelberg, Germany
e-mail: jan.schnellenbach@awi.uni-heidelberg.de

T. Baskaran
Georg-August-Universität Göttingen, Wirtschaftswissenschaftliche Fakultät, Platz der Göttinger Sieben 3, 37037 Göttingen, Germany
e-mail: tbaskar@uni-goettingen.de

G. Buenstorf et al. (eds.), *The Two Sides of Innovation*, Economic Complexity and Evolution, DOI 10.1007/978-3-319-01496-8_10,

the process of innovation. These contributions are often concerned with the immediate effects of certain institutions and policies on innovative activities in the private sector, and they come to very important insights. One can, however, also push this approach a little further and ask which institutions of government lead to the adoption of policies that in turn create a more or less friendly framework for private sector innovation. In this paper, we provide some evidence on the speed of structural change under different fiscal constitutions.

We concern ourselves with periods of intense structural change, which can be seen as an aggregate-level manifestation of processes of creative destruction. Old industries or even entire sectors decline, factors of production become temporarily abundant and eventually migrate into other sectors of the economy. Often, they are absorbed by new industries or services, which are the result of innovative entrepreneurship. The decline of old and the rise of new industries thus go hand in hand. While this may be a trivial insight from a purely economic point of view, one should remember that times of intense structural change are also accompanied by high costs of adjustment, by uncertainty and thus by challenges for economic policy-makers.

In this paper, we investigate whether different fiscal institutions lead to a variation in policy-makers' responses to the decline of ripe industries. In particular, we are interested in the effects of fiscal autonomy and competition: Are regions that are subjected to fiscal competition more or less inclined to support processes of structural change than regions in a unitary state, or regions in a system of cooperative federalism? To shed some light on this question, we look at three neighboring regions with a broadly similar industrial history: the *Saarland*, *Lorraine* and *Luxembourg*. The argument will proceed as follows: In Section 2, we give a very brief overview over issues related to structural change and develop the hypotheses to be tested. In Section 3, we offer some background information on the regions studied in our empirical analyses. Section 4 contains the econometric analysis and Section 5 concludes.

2 The Political Economy of Structural Change

2.1 Economic Policy and Structural Change

The case study we are conducting in this paper is focused on long-term structural change: the decline of the steel and mining industries in the *Saarland*, *Lorraine* and *Luxembourg*. In this sense, we are particularly interested in the destructive side of creative destruction, i.e. the fading of old industries that necessarily precedes or accompanies the emergence of novel uses of factors of production. The focus of this study also leads us to be not so much interested in the micro-foundations of innovative creation of new industries in structural change, but in the effect of fiscal institutions on the political response to sectoral decline. Much research has been conducted in innovation economics in order to understand thoroughly entrepreneurial activities that are at the center of the famous Schumpeter (1942) notion of *creative destruction*. Through analyzing the decline of an old industry in this paper, we believe to be contributing to the understanding of creative destruction from a different angle.

If economic policy aims at stabilizing the old industries, the emergence of new economic structures may be obstructed. This can be due to resources remaining employed in old sectors, or it may be due to competitive advantages that incumbent industries hold solely due to support from the public sector. In any case, the dynamics of innovative entrepreneurial activity benefits from an economic policy that allows or even facilitates decline, and such activities are slowed down by interventions favoring established industries.[1] By analyzing the impact of fiscal institutions on the propensity of governments to act in order to preserve the status quo, we thus also contribute to the understanding of institutional influences on long-term economic development through creative destruction.

There is a growing literature that shows that governments can and do slow down processes of structural change. Examples are the reluctance to quickly adopt new technologies (Fagerberg et al. 1999), the failure to foster investment in human capital that is conducive to innovation (Aghion and Howitt 2006), and the failure to facilitate intersectoral mobility in the labor market.[2] The flipside of these considerations is that governments can also successfully facilitate structural change, either through a reversal of the aforementioned policies, or through the support of new sectors (see e.g. United Nations 2006), or even through the support of sub-central governments with the means for an active structural policy (e.g. Beugelsdijk and Eijffinger 2005 with an analysis of EU structural funds).

While the technical ability to accelerate structural change may be there, political economy arguments suggest that a tendency instead to preserve the status quo is the likely outcome. Interest group theory implies that beneficiaries of structural preservation can organize their political pressure relatively easily (Olson 1965). Careful analysis of actual processes of economic policy-making reveals a widespread propensity to prefer a familiar status quo to an uncertain future (Meier and Durrer 1992). Theories of political communication under rational ignorance show that equilibria may occur where efficient policies are simply not accepted by a majority of individuals (e.g. Schnellenbach 2005 or Caplan 2007). The interesting question is thus whether different degrees of fiscal autonomy can be expected to reduce this status quo bias.

[1] See e.g. Grossman and Helpman (1991) as well as Aghion and Howitt (1998, 2006).

[2] See e.g. Caballero et al. (2004), Lee and Wolpin (2006) and Wacziarg and Wallack (2004).

2.2 Fiscal Federalism and Structural Change

Systems of cooperative federalism are typically characterized by far-reaching vertical and horizontal fiscal equalization schemes. These have several different effects. Fiscal transfers in conjunction with federal social security systems have been shown to absorb almost completely the negative shocks on household incomes (Büttner 2002; Asdrubali et al. 1996), which is likely to reduce the urge to advance structural change on the sub-central level. Politically, this implies that a single sub-central government is not fiscally punished to any great extent for implementing inefficient policies. We therefore expect the status quo bias sketched above to be dominant in cooperative federalism.

Hypothesis 1. *Transfers in cooperative federalism are not used to accelerate structural change.*

Under competitive federalism, on the other hand, far-reaching tax and spending autonomy and the absence of extensive fiscal equalization programs prevail. Regions cannot externalize the costs of structural preservation to other regions, and regional income shocks are cushioned to a lesser extent. Incumbent politicians can be held accountable through yardstick competition for bad policies (Besley and Case 1995). We therefore expect regions with high fiscal autonomy to be characterized by a lesser tendency to engage in structure-preserving policies.

Hypothesis 2. *Transfers to regions within a system of cooperative federalism will lead to a slower decline in employment in a declining sector relative to a region in a system of competitive federalism.*

In unitary states, there are often no formal fiscal transfers to regions suffering from incumbent industries in structural decline. Instead, it is common practice to lobby directly at the central-government level for support. There, the costs of supporting a single region hit by an asymmetric shock may easily appear small relative to the federal budget, which increases the probability of successful lobbying. Furthermore, there is a knowledge problem: Knowledge of efficient regional structural policies may be difficult to centralize, and political economy arguments even suggest that regional representatives have an incentive to convey false information to the central level (Besley and Coate 2003).

Hypothesis 3. *Regions in a system of cooperative federalism manage structural change at a faster pace compared to regions in a unitary state.*

Before we test these hypotheses econometrically in Section 4, we present some background information on the regions investigated in this study in the following Section 3.

3 Saarland, Lorraine and Luxembourg: Some Background Information

We conduct a case study of the three regions, *Lorraine, Luxembourg* and *Saarland*, which are characterized by similar structural conditions and developments through much of their modern economic history, but by differing institutional frameworks through the entire post-war period. We will start by looking at long-run structural change since the 19th century. The main aim of this first step of our analysis in this section is to lay out to what extent the economic structures of these three regions are similar enough for them to be suitable for our case study. In Section 4, deviations in regional development in the post-war period between the *Saarland* and *Lorraine* and *Luxembourg* are explained by the extent to which the *Saarland* has been embedded in the German fiscal equalization system. This is done by means of a formal econometric analysis.

3.1 The Institutional Framework

The three regions the long-term development of which we investigate can each serve as a proxy for one of the ideal types of federal fiscal institutions discussed above. *Lorraine* is a region in North-Eastern France and shares a border with Germany. It currently has a population of just over 2.3 million individuals and is comprised of four *Départements*. In our study, it represents a region in a unitary state. It should be noted that France has experienced a certain extent of decentralization since 1982, but remains a highly centralized county where regions have only very limited competencies. It thus remains reasonably close to the ideal type of a unitary country.

The *Saarland* is a German *Land* which is adjacent both to *Lorraine* and to *Luxembourg* (Fig. 1). It is home to approximately one million citizens, and it represents a sub-central entity in a regime of cooperative federalism in this study. Compared to the French institutional regime, German federalism allows for a considerably larger scope of autonomous decision-making for its *Länder* on the expenditure side of the budget.[3] At the same time, they have no autonomy to set rates for any of the important taxes. The lack of tax autonomy, and fiscal competition in particular, distinguishes the German regime from competitive federalism.

Finally, *Luxembourg* serves as a proxy for the small, autonomous region. Indeed, it is an autonomous country with nearly half a million citizens. Apart from transfers from the European Union, where it is subject to the same general rules as other countries and regions (and where it traditionally serves as a net contributor), *Luxembourg* has no hope to receive grants-in-aid in case of structural slumps.

[3] As Büttner et al. (2004) show, this does indeed result in measurable heterogeneity in public spending between the *Länder* in Germany.

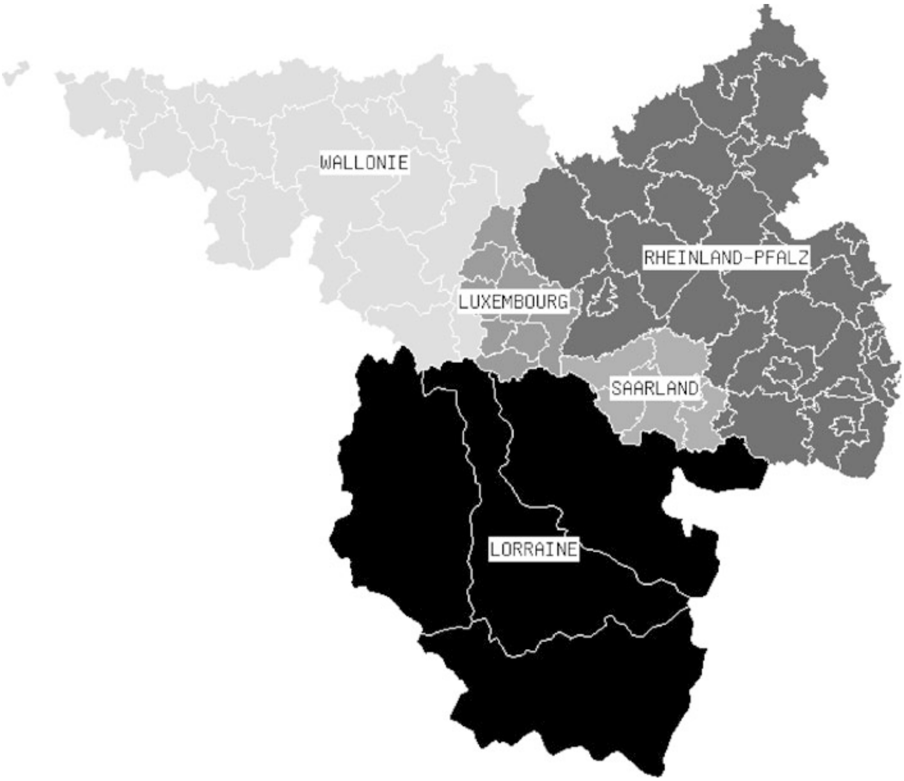


Fig. 1 A map of Saarland, Lorraine and Luxembourg with the neighboring regions Rheinland-Pfalz (Germany) and Wallonie (Belgium). Source: Geographisches Informationssystem der Großregion GIS-GR, http://www.granderegion.net/de/grande-region/kartographie/index.html

It is, therefore, reasonably safe to assume that the burden of regional structural change has been carried by the country itself. This, however, is exactly the crucial characteristic of a region in competitive federalism: It has autonomy over taxing and spending, and it cannot externalize costs of structural preservation to other jurisdictions—neither vertically to a central government, nor horizontally to other sub-central jurisdictions. For the purpose of our study, *Luxembourg* thus comes very close to a region embedded into an ideal type of a competitive federalist constitution.

3.2 Regional Economic Development in the Three Regions

The origins of the expansion of the steel and mining industry differ between the regions discussed here. For the *Saarland*, the abolishment of tariffs within Germany has opened new markets (see Banken 2002b), and industrialization in Germany in conjunction with the rapid expansion of the railroad network sharply increased demand for its coal and ore. This led to a faster paced industrialization in the *Saarland* relative to the other two regions (Banken 2002a). In particular, *Luxembourg* industrialized late.[4] An important reason for the late industrialization of *Luxembourg* has been the need for a technological innovation: Only in 1879 has the so-called *Thomas-Technique* been invented, which finally allowed for the processing of phosphor-rich *Minette* ore, found in the mines in southern *Luxembourg*, at lower costs. In the subsequent years, the steel and mining sector boomed, financed to a great extent with capital inflows from Germany. Consequently, *Luxembourg* industrialized quickly, and the relative weight of the steel and mining industries in *Luxembourg* arrived at about the same size as those in *Saarland* and *Lorraine*.

There also has been a high degree of economic integration in terms of cross-ownership between our three regions before World War I, as documented by Banken (1995, 2000: 257, 2002c, 2003: 310). While the two world wars did have a temporary impact on industrial structure, and even though *Luxembourg* initiated the next step of structural change relatively early,[5] structural convergence between the three regions in terms of the relative importance of the steel and mining industries has been robust. As Fig. 2 shows, the employment shares of steel and mining relative in total employment are relatively similar, even in 1963, at the beginning of our time series.

Towards the end of the 1950s, post-war reconstruction began to slow down and the Korean War had ended. Accordingly, demand for steel was subject to a short-term decline. On the supply side of the market, new competition from the developing countries in Asia and Australia began to emerge[6] and market shares of European producers declined. Technological innovations led to an increase of capital intensity in the steel and mining industries, but also to changes in the demand for coal and steel from the big industrial nations. Consequently, these sectors have absorbed ever smaller fractions of the labour force, as illustrated in Fig. 2.

During both oil price shocks, the structural decline of the coal and steel industries was briefly interrupted because coal could substitute for oil. This was particularly notable at the beginning of the 1980s. This interruption was, however, much less important in steel demand, as the series for *Luxembourg* and *Lorraine* show, which produce iron ore and steel to a relatively larger extent than the *Saarland*.

[4] In 1871, around 60% of the population were still employed in the primary sector. Industrial production was mostly confined to small companies, which pioneered in establishing the seeds of a steel and mining industry (Pohl 1999).

[5] The establishment of *Luxembourg* as a financial center was initiated in the interwar period. See Anders (1932).

[6] Australian coal exports increased from 2 million tons of coal equivalents (MTCE) on 1960–64 to 106 MTCE at the beginning of the 1990s (Miehe-Nordmeyer et al. 1998: 538). On the demand side, despite increasing GDP, a decline of coal consumption in Western Europe from 469 MTCE in the years 1960–64 to 408 MTCE in 1990–92 occurred (Miehe-Nordmeyer et al. 1998: 539).

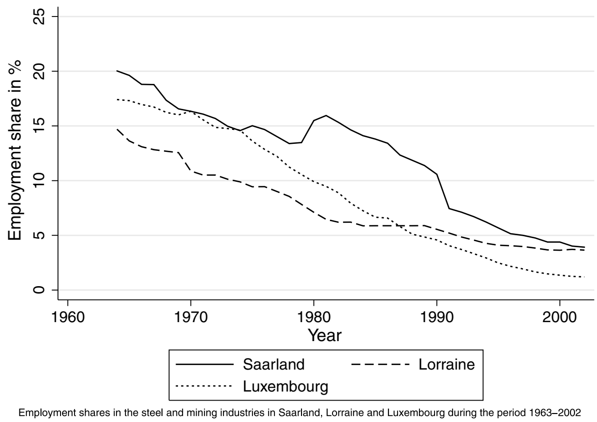


Fig. 2 Employment shares in the steel and mining industries in the *Saarland*, *Lorraine*, and *Luxembourg* during the period 1963–2002. Data sources: Statistical Office of the *Saarland* (Employment share of steel and mining in the *Saarland*), INSEE (Employment share of steel and mining in *Lorraine*), Statec Luxembourg (Employment share of steel and mining in *Luxembourg*). For further details, please see footnote 8

Increased energy prices after the oil price shocks also induced a reduction of steel demand, for example in the automotive industries (Miehe-Nordmeyer et al. 1998: 539). The policy response to the decline of coal and steel industries adopted by EU member countries, coordinated in the European Coal and Steel community, was a strategy of price subsidization in order to keep the coal and steel industries artificially competitive. In addition, the *Saarland* subsidized its heavy industries directly and was in part able to finance this through transfers received from the fiscal equalization system. This is corroborated by several government reports, e.g. Sievert and Streit (1964). In 1969, after experiencing a first crisis with a decline in the coal industry, Biehl et al. (1969) cautiously discussed the alternative strategy of actively pushing the *Saarland* towards alternative industries by driving back the coal and steel industries to less importance.

After the crisis of the steel industry in the 1970s, Augustin et al. (1978) suggested to promote other industries more heavily in an effort to speed up structural change. However, in 1982, Lenhof et al. (1982) interpreted the still continuing subsidization of the coal and steel industries as a strategy of social policy conducted in order to counteract the income effects of structural change. At the beginning of the 1990s, Bade (1990) and Sievert et al. (1991) illustrated that the *Saarland* was not successful in achieving structural change and that wages and salaries had to adjust in order to raise the attractiveness of the region for investors from other industries. The bankruptcy of *Saarstahl* at the beginning of the 1990s illustrated the validity of their verdict.

Indeed, there is some evidence that wages did not adjust sufficiently to induce structural change. From the beginning of the 1970s until the beginning of the 1990s, real wages in *Saarland* relative to Germany as a whole increased. According to the studies by Miehe-Nordmeyer et al. (1998) and Miehe-Nordmeyer (2001), the relative growth of employment in the *Saarland* was not significantly affected by relative growth in wages, and vice versa. The decline of heavy industries did thus not induce the expected negative effect on wages, but only led to higher unemployment.

Regarding *Lorraine*, the development was similar, although the French government started to restructure the economy by closing plants more directly and earlier compared to the *Saarland*. In the case of *Luxembourg*, the shift towards service industries, also due to a liberal financial market regulation, happened earlier and much more successfully (Benini 1996; Miehe-Nordmeyer 2001). Finally, at the beginning of the new millennium, the importance of the coal and steel industries as a share of total employment converged somewhat between the three regions, as Fig. 2 shows.

4 Structural Change in the Post War Period: a Time-series Analysis

4.1 The Empirical Strategy

The goal of this section is to explore whether the degree of fiscal autonomy of a region is related to the speed with which it embraces and fosters structural change. In order to explore this question, we will focus on the role of intergovernmental transfers. The reason we focus on intergovernmental transfers is that their importance is a good proxy for the extent of fiscal autonomy that sub-central entities possess within a federation. It is a peculiarity of sub-central entities in cooperative federalism that they can become recipients of vertical and horizontal fiscal transfers from other jurisdictions, while having significant spending autonomy. In a unitary state, there may also be transfers of resources to regions, but they are not fiscal transfers to sub-central entities with spending autonomy. In this sense, fiscal transfers to a region in cooperative federalism increase the spending autonomy of this region. Given our arguments above, we expect the increased spending autonomy *ceteris paribus* to be associated with an increased speed of structural change.

However, an increase in transfers is also associated with a loss of autonomy on the revenue side of the budget. Sub-central entities become less dependent on own sources of revenue, and therefore less interested in fostering their own tax bases. Given our arguments above, we expect the loss of autonomy on the revenue side following from increased transfers to be *ceteris paribus* associated with a decreased speed of structural change.

In order to address the question as to the sign of the overall effect of intergovernmental transfers on the speed of structural change, we pursue the following empirical strategy. Our approach is based on the argument that the *Saarland*, *Lorraine*, and *Luxembourg* possess similar economic structures but face distinct fiscal regimes. That is, the economies of all three regions have been historically dominated by the mining and steel industries. This suggests that all three regions have been similarly affected by the changing economic environment after World War II. On the other hand, the three regions have operated, since World War II, under different fiscal regimes. The *Saarland* has been part of the German Federation, *Lorraine* part of unitary France, and *Luxembourg* an independent country. Consequently, it is a reasonable conjecture that relative differences in the speed of structural change in these three regions are, at least in part, due to the differences in their fiscal regimes.

A straightforward empirical approach to investigate the links between transfers and structural change is to look at the employment in the steel and mining industries for the *Saarland* relative to employment in each of the other two regions, and to investigate whether relative employment is affected by transfers. The reason for our focus on the relative employment share in the *Saarland* relative to the other regions is that reliable and consistent data on transfers are available for the *Saarland*, due to the existence of a formal equalization scheme in the German federation, but not for *Lorraine*. *Luxembourg*, on the other hand, is an independent country and did not receive any transfers.

In the empirical investigation, we therefore focus on the relationship between transfers and structural change in the *Saarland*, relative to *Lorraine* and to *Luxembourg*. Data on equalization transfers paid to the the *Saarland* during this period were obtained from the German Federal Statistical Office.[7] These data consist of the sum of horizontal (*Länderfinanzausgleichszahlungen*) and vertical (*Bundesergänzungszuweisungen*) transfers per capita. Data on employment in the steel and mining sectors for the *Saarland* are obtained from the Statistical Office of the *Saarland*. Data on employment in the steel and mining industries in Lorraine are obtained from the INSEE (Institut national de la statistique et des Études Économiques). Finally, data on employment in this sector in Luxembourg are obtained from Statec Luxembourg.[8] Summary statistics are reported in Table 1. Our data on transfers and the steel and mining employment shares in the *Saarland* covers the period 1961–2004. The data on the steel and mining employment shares

[7] These data are also available publicly from the website of the Federal Finance Ministry at http://www.bundesfinanzministerium.de/nn_3264/DE/Wirtschaft__und__Verwaltung/Finanz__und__Wirtschaftspolitik/Foederale__Finanzbeziehungen/Laenderfinanzausgleich/node.html?__nnn=true. The Federal Statistical Office kindly provided us with a complete datafile of transfers.

[8] The employment data for the Saarland (employment share of steel and mining in the Saarland) are obtained from the Statistical Office of the Saarland, provided by Karl Schneider, who is member of the working group of statistical offices of the greater region Saarland-Lorraine-Luxembourg-Rheinland-Pfalz-Wallonie. The data for Lorraine are provided by INSEE (employment share of steel and mining in Lorraine) and the data for Luxembourg are provided by Statec Luxembourg (employment share of steel and mining in Luxembourg) to this working group. In addition to these sources, missing observations could be inferred from Augustin et al. (1978), Lenhof et al. (1982), Sievert et al. (1991) and Miehe-Nordmeyer (2001). The transfer data are obtained from the German Federal Statistical Office (Transfers per capita), see footnote 7.

Table 1 Summary statistics

Variable	Mean	Std. dev.	Min.	Max.	N
Transfers per capita to the *Saarland*, nominal values (1961–2004)	353.229	336.93	60.018	1125.253	44
Employment share of steel and mining in the *Saarland* (1961–2004)	12.397	5.615	3.73	22.06	44
Employment share of steel and mining in *Lorraine* (1963–2004)	7.522	3.621	3.42	16.84	42
Employment share of steel and mining in *Luxembourg* (1961–2002)	9.406	5.976	1.2	18.02	42

Sources: German Federal Statistical Office (Transfers per capita to the *Saarland*), Statistical Office of the *Saarland* (Employment share of steel and mining in the *Saarland*), INSEE (Employment share of steel and mining in *Lorraine*), Statec Luxembourg (Employment share of steel and mining in *Luxembourg*). For further details, please see footnote 8

in Lorraine are available for the period 1963–2004, while the data on the employment shares in Luxembourg are available for the period 1961–2002.

Conceptually, taking *Luxembourg* as the point of reference, our approach entails that we look at the employment shares of the steel and mining industries relative to total employment of the *Saarland*, divided by that of *Luxembourg*. A positive impact of transfers to the *Saarland* on this dependent variable would mean that transfers led to a slower decline of employment in the ailing sector in *Saarland* relative to *Luxembourg*. A negative impact, on the other hand, would imply that transfers were used to foster structural change, and that labor migrates to other industries faster in the *Saarland* than in Luxembourg because of these transfers.

Similarly, taking *Lorraine* as the point of reference (i.e., looking at the series of relative employment shares in *Saarland* and *Lorraine*), we can investigate whether in a system of cooperative federalism, subnational governments combine transfers with local knowledge to foster structural change more effectively than in a unitary regime that has to cope with the problem of centralizing policy-relevant knowledge. Again, a negative effect of transfers would indicate that this is the case, while a positive effect would indicate that even vis-à-vis unitary regimes, structural change is slowed down in a system of cooperative federalism.

We therefore estimate two different models, one using the employment in the steel and mining industries in the *Saarland* relative to that in *Luxembourg*, and one relative to that in *Lorraine*. From our earlier theoretical considerations, it follows that we expect different signs on the coefficients of the transfer variable for the two different types of regimes. If the institutional feature of higher sub-central spending autonomy in cooperative federalism compared to unitary regimes is indeed important in order to efficiently exploit local knowledge, then increased fiscal transfers to the *Saarland* should increase the speed of structural change, relative to unitary *Lorraine*, since sub-central spending autonomy is the crucial institutional feature distinguishing these two regions. The crucial institutional difference between the *Saarland* and *Luxembourg*, on the other hand, is the lack on revenue autonomy of the former. Here, we would therefore expect the negative effect of transfers on the relative speed of structural change to be dominant.

4.2 Estimation and Results

A necessary requirement for the empirical analysis is that the data used in the estimations are stationary. It is well known that non-stationarity could lead to spurious regression results. We have employed a standard set of diagnostic tests in order to check for non-stationarity in our variables. First, the autocorrelation functions of both the relative employment series and the transfer series for the *Saarland* indicate autoregressive processes. This in turn suggests the presence of unit roots in these series.

Given the findings from the autocorrelation functions, we subsequently employed unit root tests for the levels of all variables used in the analysis. The tests regarding the transfer series were conducted for the log of the real values instead of the nominal values in levels. Using real values is obviously more appropriate than nominal values for our empirical analysis. Similarly, log-transforming the (real) transfers is appropriate since percentage changes rather than absolute changes in transfers are likely to be related to employment shares. The results do not allow us to reject the hypothesis of the presence of a unit root even at the 10% confidence level for the level series.[9]

We have also conducted the Johansen rank test for cointegration. However, no evidence for cointegration was found, which precludes us from employing an error correction model on the level series. We therefore turned to the first difference series, which implies that we look at periodic changes in transfers and employment shares, rather than levels. The unit root tests for the first differences reject the hypothesis of a presence of a unit root at the 1% critical level.[10] Overall, these results suggest that the level series are first-difference stationary and not cointegrated.[11] We can therefore apply standard time-series econometric tools to the first difference series, and estimate ADL as well as VAR models.

After establishing that we can conduct the empirical analysis using the first differenced data, we proceed to specify the empirical model. A major challenge in exploring the relationship between transfers and structural change empirically is

[9] The Augmented Dickey-Fuller (ADF) test for the Saarland relative to Lorraine level series produced a test statistic of −1.2296, with critical values of −3.6056 (1%-level), −2.9369 (5%-level) and −2.6069 (10%-level) and the lag order suggested by the Schwartz Information Criterion (SIC). For Saarland relative to Luxembourg, the ADF test statistic is 1.3537 with critical levels of −3.601, −2.935 and −2.6058, respectively.

[10] The ADF test statistic for Saarland relative to Lorraine is −4.4287 with a critical value of −3.6056 on the 1%-level. For Saarland relative to Luxembourg, the ADF test statistic is −4.6381 with a critical value on the 1%-level of −3.6056.

[11] All test results were calculated with EViews.

that, while transfers might affect structural change, structural change might also affect the amount of transfers a region receives. For example, the declining industrial base and soaring unemployment in the *Saarland* during the time-frame of the analysis might have led the federal government to grant additional transfers to the *Saarland*. If the possibility that structural change increases government transfers is not explicitly recognized and sufficiently addressed in the estimations, we may reach invalid conclusions. Consequently, we take this problem into account in our empirical approach, and estimate models to explore how transfers affect structural change *and* how structural change affects transfers.

In a first step, we thus estimate two ADL-models in first differences for each pair of regions, namely

$$\Delta E_t = \alpha + \sum_{k=1}^{K} \beta_k \Delta E_{t-k} + \sum_{k=0}^{K} \delta_k \Delta T_{t-k} + \sum_{n=0}^{N} \gamma_n D_n + \varepsilon_t \tag{1}$$

and

$$\Delta T_t = \phi + \sum_{k=1}^{K} \tau_k \Delta T_{t-k} + \sum_{k=0}^{K} \kappa_k \Delta E_{t-k} + \sum_{n=0}^{N} \rho_n D_n + \mu_t \tag{2}$$

where ΔE_t and ΔT_t denote the first differences of relative employment shares and of the log of (real) transfers per capita at time t, and k denotes the lag. We include up to three lags of both the employment share and transfers variables to enable the models to capture long-run effects. D denotes a vector of additional variables. This vector consists of a trend for the 1994–2002 period (*Smalltrend*) to take account of the fact that the *Saarland* had benefited from extraordinary federal bailouts during this period, and a dummy for 1991 (*Break*), where the steel industry of the *Saarland* was hit by an extremely large negative shock. The dummy variable is thus expected to have a negative coefficient when regressed on the relative employment share of the Saarland.[12]

In the ADL-regressions, the Durbin-h autocorrelation test suggests autocorrelation for the employment share model (Eq. 1) when employment relative to *Lorraine* is used as dependent variable. The Durbin-h test statistic for this case is 0.63 with a p-value of 0.04. When employment relative to *Luxembourg* is used, no evidence for autocorrelation is found. The Durbin-h test statistic is 0.13 with a p-value of 0.63. Similarly, no autocorrelation is found for both transfer models (Eq. 2). The Durbin-h test statistics are 0.58 and 0.16, respectively, and the p-values are 0.18 and 0.65.

Thus, the H0 of zero autocorrelation is rejected at the 10% level for the employment model relative to *Lorraine*. Typically, if autocorrelation is found, further lags have to be included until the autocorrelation test is rejected. However, since the number of observations is rather limited in our case, including additional

[12] All regressions were run with Stata.

Table 2 ADL models

	Lorraine	Luxembourg
DEmployment(−1)	0.294* (0.149)	0.096 (0.127)
DEmployment(−2)	0.034 (0.164)	−0.058 (0.139)
DEmployment(−3)	−0.190 (0.165)	0.203 (0.159)
DTransfers	−0.154 (0.101)	0.130 (0.082)
DTransfers(−1)	0.012 (0.092)	0.055 (0.070)
DTransfers(−2)	−0.033 (0.090)	0.031 (0.070)
DTransfers(−3)	−0.037 (0.090)	0.095 (0.070)
Break	−0.501*** (0.119)	−0.559*** (0.096)
Smalltrend	−0.008 (0.007)	0.010 (0.008)
Constant	0.025 (0.024)	0.037* (0.019)
N	38	38
F	2.802	4.994
Adj. R2	0.305	0.493
Root MSE	0.116	0.090
Degrees of freedom	28	28
Joint significance of transfers (F-test, p-value in parentheses)	0.61 (0.66)	1.12 (0.37)

Dependent variable: Employment share of the *Saarland* steel and mining industry relative to Lorraine and Luxembourg
Standard errors in parentheses; *, ** and *** denote significance on the 10, 5 and 1 percent level
Sources: Own calculations based on the data sources mentioned in Table 1
The estimation covers the period 1963–2004 for *Lorraine* and the period 1961–2002 for *Luxembourg*

lags might reduce the degrees of freedom to such an extent that the coefficent estimates are no longer trustworthy. (If too many parameters have to be estimated with a limited number of observations, the estimates will be characterized by high variability.) Moreover, autocorrelation tests conducted after the VAR-estimations reported further below are never rejected[13] and thus do not suggest autocorrelation in any of these models. For these two reasons, we limit our lag length to three in all our regressions. However, we have also estimated all regressions with up to five lags included. These results, which are not reported but available upon request, confirm the results and interpretations from models with up to three lags.

The results for the ADL-models are presented in Tables 2 and 3. In the first table, the first difference of the relative employment share of the *Saarland*, relative to *Lorraine* (first column) and *Luxembourg* (second column), is the dependent variable. For *Luxembourg*, we see that only the structural break dummy has a significant impact, with the expected sign. All other explanatory variables have insignificant coefficients. For *Lorraine*, we observe that the first differences of employment

[13] We applied the LM-test for the VAR estimations. The test-statistics for the models relative to *Lorraine* and *Luxembourg* were 4.76 and 4.29, respectively, with p-values of 0.31 and 0.37 even for two lags.

Table 3 ADL models

	Lorraine b/se	Luxembourg b/se
DEmployment	−0.495 (0.326)	0.633 (0.399)
DEmployment(−1)	−0.067 (0.285)	−0.007 (0.283)
DEmployment(−2)	0.493* (0.279)	0.421 (0.298)
DEmployment(−3)	−0.731** (0.270)	−1.033*** (0.303)
DTransfers(−1)	−0.016 (0.165)	−0.071 (0.156)
DTransfers(−2)	−0.085 (0.161)	−0.120 (0.154)
DTransfers(−3)	−0.143 (0.160)	−0.105 (0.158)
Break	−0.291 (0.268)	0.494 (0.301)
Smalltrend	−0.026** (0.012)	−0.017 (0.017)
Constant	0.069 (0.042)	0.034 (0.044)
N	38	38
F	1.576	1.878
Adj. R2	0.123	0.176
Root MSE	0.208	0.199
Degrees of freedom	28	28
Joint significance of employment (F-test, p-value in parentheses)	2.63 (0.06)	3.51 (0.02)

Dependent variable: Transfers received by the *Saarland*. Employment variables: Employment share of the Saarland steel and mining industry relative to Lorraine and Luxembourg
Standard errors in parentheses; *, ** and *** denote significance on the 10, 5 and 1 percent level
Sources: Own calculations based on the data sources mentioned in Table 1
The estimation covers the period 1963–2004 for *Lorraine* and the period 1961–2002 for *Luxembourg*

shares appear to be autoregressive, albeit with a weakly significant coefficient. Apart from the first lag of the employment share, only the structural break dummy is significant with the expected sign in the *Lorraine* model.

Regarding our variables of interest, the transfers to the *Saarland*, the results in Table 2 for both the *Luxembourg* and *Lorraine* models obviously suggest that they have no effect on employment.

In Table 3, we find that the third lag of the employment share has a highly significant negative effect on transfers in both the *Lorraine* and the *Luxembourg* model (while the second lag has a positive but only weakly significant effect on transfers in the *Lorraine* model). A reasonable interpretation for this result is that a decline in employment shares leads to an increase of transfers with a delay of about three years in the *Saarland*.

As a first step to validate these results, we performed the joint F-Test on the explanatory variables in both models. The results reported in Tables 2 and 3 support, irrespective of whether *Lorraine* or *Luxembourg* is used as the reference, the hypothesis that employment affects transfers (p-values for the F-statistics of 0.06 and 0.02, respectively), but not the opposite hypothesis (p-values for the

F-statistics of 0.66 and 0.37, respectively). This again suggests the interpretation that changes in employment drive changes in transfers and not vice versa.[14]

However, one underlying problem with ADL models is that they rest on a strict exogeneity assumption concerning the right hand side variables, which for theoretical reasons is not guaranteed in our case, with a plausible circular effect between transfers and employment. We have, therefore, as the second step in the empirical analysis and as a robustness check, specified a VAR model. The advantage of a VAR model is that all dependent variables can be treated as endogenous. It is constructed on the basis of Eqs. 1 and 2, in the sense that these equations are now assumed to describe a system. The difference between the VAR models and the ADL-models specified in Eqs. 1 and 2 is that, in the VAR, the contemporaneous values of transfers and employment shares are omitted as control variables because both are treated as endogenous. Instead, ΔE_t and ΔT_t are explained with their lagged values, i. e. in terms of their common history.

Table 4 reports the results for the employment model using the VAR approach. As in the respective ADL model, the dependent variable is the first difference of relative employment shares between either the *Saarland* and *Lorraine*, or the *Saarland* and *Luxembourg*.

The results show that lagged values of transfers have neither individually nor jointly a significant effect on employment. That is, the individual coefficients are all insignificant and the F-test on joint significance of the transfers in the employment model, reported at the bottom of Table 4, is also insignificant, irrespective of whether *Lorraine* or *Luxembourg* is used as the benchmark (with p-values of 0.99 and 0.60, respectively).

The results for the VAR-estimation of the transfer model are reported in Table 5. We find that the third lag of relative employment shares has a negative coefficient and is significant at the 1% level. That is, we find, as in the ADL models, that with a lag of three years, a (relative) decline in employment in the *Saarland* leads to a significant increase of transfers. Moreover, the lags of the employment shares in the transfer model are jointly significant irrespective of whether *Lorraine* or *Luxembourg* is used as the benchmark (with a p-value of 0.05 and 0.01, respectively.)

This supports the conjecture that transfers do not affect relative employment shares, but that a decline in employment leads to higher transfers to the *Saarland* with a lag of three years.

[14] We also conducted Jarque–Bera Normality tests after our regressions. We found that the years 1970, 1980, and 1994 were potential outliers. Therefore, we also conducted all regressions with dummies for these particular years included. We found, as in the paper, that only employment shares had a significant or almost significant joint effect on transfers while transfers were jointly insignificant for employment shares. The individual estimates were somewhat different, however. When these dummies are included, the first lag of employment shares is negative and significant for transfers, while the third lag is significantly positive. Overall, these results suggest that transfers react somewhat faster to structural change than those reported in the paper.

Table 4 VAR models

	Lorraine b/se	Luxembourg b/se
DEmployment(−1)	0.329** (0.150)	0.104 (0.130)
DEmployment(−2)	−0.045 (0.159)	−0.004 (0.139)
DEmployment(−3)	−0.084 (0.153)	0.075 (0.140)
DTransfers(−1)	0.016 (0.094)	0.050 (0.072)
DTransfers(−2)	−0.021 (0.092)	0.017 (0.071)
DTransfers(−3)	−0.016 (0.091)	0.089 (0.071)
Break	−0.494*** (0.122)	−0.539*** (0.098)
Smalltrend	−0.005 (0.006)	0.009 (0.008)
Constant	0.015 (0.023)	0.045** (0.018)
N	38	38
Log-L	43.804	54.750
AIC	−1.358	−1.934
Degrees of freedom	28	28
Joint significance of transfers (F-test, p-value in parentheses)	0.12 (0.99)	1.89 (0.60)

Dependent variable: Employment share of the *Saarland* steel and mining industry relative to Lorraine and Luxembourg

Standard errors in parentheses; *, ** and *** denote significance on the 10, 5 and 1 percent level

Sources: Own calculations based on the data sources mentioned in Table 1

The estimation covers the period 1963–2004 for *Lorraine* and the period 1961–2002 for *Luxembourg*

4.3 Discussion

The evidence we have reported above supports our Hypothesis 1, since it indicates that fiscal transfers in regimes of cooperative federalism are not used to accelerate structural adjustments. Transfers in a federal system do not lead to a swifter decline of employment in an ailing industry; rather, they are caused by the regional decline of an industry and increase with its severity. This is in sharp contrast to statements that are frequently made in the political sphere in order to justify fiscal transfers to ailing regions. The most important argument made in this context is that transfers endow the receiving regions with the resources necessary to manage structural adjustments themselves and to speed up such adjustments. In light of our evidence, this does not appear to be the case. Instead, the evidence is consistent with theoretical approaches that predict a status quo bias in structural policies.

Interestingly, the results do, however, also not show a significant slowdown of structural decline in the *Saarland* relative to the other two regions. In particular, transfers received by the *Saarland* do not have a statistically significant effect on its steel and mining employment shares relative to those of *Luxembourg*. We, therefore, do not find supporting evidence for our Hypothesis 2—being exposed to structural decline as a small open economy without resort to transfers does not lead to a significantly faster decline of employment in the declining sector.

Table 5 VAR models

	Lorraine b/se	Luxembourg b/se
DEmployment(−1)	−0.230 (0.270)	0.058 (0.287)
DEmployment(−2)	0.515* (0.285)	0.419 (0.306)
DEmployment(−3)	−0.689** (0.275)	−0.986*** (0.310)
DTransfers(−1)	−0.024 (0.169)	−0.040 (0.159)
DTransfers(−2)	−0.074 (0.165)	−0.110 (0.157)
DTransfers(−3)	−0.135 (0.164)	−0.049 (0.157)
Break	−0.046 (0.219)	0.153 (0.216)
Smalltrend	−0.023** (0.012)	−0.012 (0.017)
Constant	0.061 (0.042)	0.062 (0.041)
N	38	38
Log-L	43.804	54.750
AIC	−1.358	−1.934
Degrees of freedom	28	28
Joint significance of employment (F-test, p-value in parentheses)	7.87 (0.05)	10.94 (0.01)

Dependent variable: Transfers received by the *Saarland*. Employment variables: Employment share of the Saarland steel and mining industry relative to Lorraine and Luxembourg
Standard errors in parentheses; *, ** and *** denote significance on the 10, 5 and 1 percent level
Sources: Own calculations based on the data sources mentioned in Table 1
The estimation covers the period 1963–2004 for *Lorraine* and the period 1961–2002 for *Luxembourg*

The finding may appear to be surprising that higher (i.e., third) order lags of the employment ratios are significant on conventional levels. Note, however, that this is not implausible, since there are often substantial implementation lags in economic policy-making. For example, a region experiencing a decline of one of its industries may receive transfers relatively quickly, but invest in the stabilization of the ailing industry only after it becomes apparent that the crisis is not temporary and that the industry will not recover when left on its own.

At the same time, the fact that the coefficients of all lagged transfer changes in Table 4 are not significantly different from zero for the regression with *Lorraine* leads us to the conclusion that we also do not find supporting evidence for our Hypothesis 3. The additional spending autonomy granted to the receiving region through fiscal transfers does not yield any detectable effect in our data. However, it is difficult to interpret further the implications of these non-significant results.[15] It appears that the variation in the data is insufficient to test this hypothesis meaningfully.

Nevertheless, there are some policy conclusions that can, with the appropriate caution, be drawn from our analysis. We have presented conclusive evidence that,

[15] We are grateful to a referee for pointing this out.

in the case study investigated here, changes in employment lead changes in transfers, and not vice versa. Often heard claims by political practitioners that fiscal transfers are used in a prudent, forward-looking fashion actively to manage and to accelerate structural change are thus not supported by the evidence. On the contrary, fiscal transfers appear to be neutral in terms of structural change. It is, of course, open to debate to what extent results from investigating transfers to a German *Bundesland* can be generalized to other cooperative federations. It is, however, a common feature of the constitutional frameworks in such federations that fiscal transfers are precipitated by regional economic decline. The conditions for truly forward-looking interventions through transfers are therefore similarly unfavorable in other federations.

This would not inhibit regional decision-makers from using transfers when they eventually flow to accelerate structural change. Again, we do not find evidence supporting such forward-looking behavior in cooperative federations. Stronger evidence with regard to the institutional differences between fiscal regimes is certainly a desideratum. However, our empirical case study is—to the best of our knowledge—a first step towards empirically establishing a link between the receipt of fiscal transfers in a cooperative federation and structural change.

5 Conclusions

In this paper, we have investigated and compared three cases of long-term interaction between fiscal institutions and structural change. From a theoretical point of view, we have argued that a status quo bias implies that political preferences in regions that are subjected to structural change typically tend towards preserving the declining incumbent industry. We have also argued that this tendency should be affected by fiscal institutions: Autonomous regions that cannot externalize costs of impeding sectoral decline are likely to be more open to aggregate processes of creative destruction.

Our empirical time-series evidence supports the central theoretical claim that fiscal transfers do not help to accelerate processes of structural change. Transfers received by a non-autonomous region are not used not foster structural change. It rather appears that sectoral decline induces transfers, which are then used to absorb negative shocks to a region in a fashion that has no direct link to structural change. The common political justifications that are usually given to vindicate such policies are therefore flawed.

Acknowledgements The authors would like to thank the German Science Foundation (DFG) for funding this project (DFG-SPP 1142). We are indebted to Heiko Burret and Fatma Deniz for assisting with some of this research. The paper has been presented at the 2009 meeting of the Evolutionary Economics Committee of the German Economic Association at Jena, at the 2009 meeting of the Public Choice Society at Las Vegas, at the 2009 meeting of the European Public Choice Society at Athens and at the 2010 annual conference of the Royal Economic Society at Guildford. We are grateful for very helpful remarks received on these occasions. Last but not least, the authors also thank two anonymous referees for considerate comments and suggestions.

References

Aghion P, Howitt P (1998) Endogenous growth theory. MIT Press, Cambridge
Aghion P, Howitt P (2006) Appropriate growth policy: a unifying framework. J Eur Econ Assoc 4:269–314
Anders J (1932) Le marché financier luxembourgeois. Office des éditions internationales, Paris
Asdrubali P, Sørensen BE, Yosha O (1996) Channels of interstate risk sharing: United States 1963–1990. Q J Econ 111:1081–1110
Augustin C, Glöckner C, Knoepfler D, Leydecker K, Sievert O, Tomann H (1978) Die wirtschaftliche und soziale Entwicklung im Grenzraum Saar-Lor-Lux. Schriftenreihe der Regionalkommission Saarland—Lothringen—Luxemburg—Rheinland-Pfalz, Saarbrücken et al.
Bade FJ (1990) Die Stellung des Saarlandes im wirtschaftlichen Strukturwandel der Bundesrepublik Deutschland. Zeitschrift für Wirtschaftsgeographie 34:195–207
Banken R (1995) Die Industriezweige der Saarregion im 19. Jahrhundert: Wirtschaftliche Interdependenzen verschiedener Branchen während der Industrialisierung. In: Leboute R, Lehners JP (eds) Passé et avenir des bassins industriel en Europe. Centre Universitaire de Luxembourg, Luxembourg, pp 39–60
Banken R (2000) Die Industrialisierung der Saarregion 1815–1914, Bd. 1: Die Frühindustrialisierung 1815–1850. Franz Steiner Verlag, Stuttgart
Banken R (2002a) Die Entwicklung des Steinkohlenbergbaus in der Saarregion 1815–1914. In: Herrmann HW, Wynants P (eds) Acht Jahrhunderte Steinkohlenbergbau. Colloques Meuse-Moselle 2. Facultés Universitaires Notre-Dame de La Paix, Namur, pp 273–290
Banken R (2002b) Die Industrialisierung der Saarregion 1815–1913. In: Pierenkemper T (ed) Die Industrialisierung europäischer Montanregionen im 19. Jahrhundert. Franz Steiner Verlag, Stuttgart, pp 59–101
Banken R (2002c) Die wachsenden wirtschaftlichen Verflechtungen der westeuropäischen Montanreviere in Frankreich, Belgien, Luxemburg und Deutschland zwischen 1890–1914. Les Cahiers Lorrains 2-2002:107–127
Banken R (2003) Die Industrialisierung der Saarregion 1815–1914, Bd. 2: Take-Off Phase und Hochindustrialisierung 1850–1914. Franz Steiner Verlag, Stuttgart
Benini R (1996) Regional development and industrial convergence policy: lessons from western European experiences. Mimeo, The World Bank, Washington, DC
Besley T, Case AC (1995) Incumbent behavior: vote seeking, tax setting and yardstick competition. Am Econ Rev 85:24–45
Besley T, Coate S (2003) Centralized versus decentralized provision of local public goods: a political economy approach. J Pub Econ 87:2611–2637
Beugelsdijk M, Eijffinger S (2005) The effectiveness of structural policy in the European Union: an empirical analysis for the EU-15 in 1995-2001. J Common Market Studies 43:37–51
Biehl D, Oldewage R, Schütz W, Sievert O, Springer P, Strassert G (1969) Strukturprogramm Saar: Möglichkeiten einer aktiven Sanierung der Saarwirtschaft. Ansatzpunkte, Maßnahmen, Kosten. Planungsgruppe beim Ministerpräsidenten des Saarlandes, Saarbrücken
Büttner T (2002) Fiscal federalism and interstate risk sharing: empirical evidence from Germany. Econ Lett 74:195–202
Büttner T, Schwager R, Stegarescu D (2004) Agglomeration, population size and the cost of providing public services: an empirical analysis for German states. Pub Fin Manage 4:496–520
Caballero RJ, Cowan KN, Engel EMRA, Micco A (2004) Effective labour market regulation and microeconomic flexibility. Cowles Foundation discussion paper 1480. Cowles Foundation, New Haven
Caplan B (2007) The myth of the rational voter. Why democracies choose bad policies. Princeton University Press, Princeton
Fagerberg J, Guerreri P, Verspagen B (eds) (1999) The economic challenge for Europe: adapting to innovation based growth. Elgar, Cheltenham

Freeman C (1995) The national system of innovation in historical perspective. Cam J Econ 19:5–24
Grossman GM, Helpman E (1991) Innovation and growth in the global economy. MIT Press, Cambridge
Lee D, Wolpin KE (2006) Intersectoral labor mobility and the growth of the service sector. Econometrica 74:1–46
Lenhof J, Lissmann J, Reichert H, Sievert O (1982) Probleme des saarländischen Arbeitsmarktes. Forschungsbericht Nr. 49. Institut für empirische Wirtschaftsforschung der Universität des Saarlandes, Saarbrücken
Meier A, Durrer K (1992) Ein kognitiv-evolutionäres Modell des wirtschaftspolitischen Prozesses. In: Witt U (ed) Studien zur evolutorischen Ökonomik II. Duncker & Humblot, Berlin, pp 229–254
Metcalfe SJ (1995) The economic foundations of technology policy: equilibrium and evolutionary perspective. In: Stoneman P (ed) Handbook of the economics of innovation and technological change. Blackwell, Oxford
Miehe-Nordmeyer G (2001) Ökonomische Anpassung an Schocks: Das Beispiel der Regionen Saarland, Lothringen und Luxemburg. Frankfurt a.M. et al: Peter Lang
Miehe-Nordmeyer G, Holzmann R, Hervé Y (1998) Asymmetrische Schocks und makroökonomische Anpassungsreaktionen im Saarland: Eine empirische Untersuchung. Zeitschrift für Wirtschafts- und Sozialwissenschaften 118:537–571
Olson M (1965) The logic of collective action. Harvard University Press, Cambridge
Pohl H (1999) Grundzüge der Wirtschaftsgeschichte Luxemburgs von der zweiten Hälfte des 19. Jahrhunderts bis in die 1920er Jahre. Vierteljahresschrift für Wirtschafts- und Sozialgeschichte 86:309–342
Schnellenbach J (2005) Model uncertainty and the rationality of economic policy. J Evol Econ 15:101–116
Schumpeter J (1942) Capitalism, socialism and democracy. Harper and Brothers, New York
Sievert O, Streit M (1964) Entwicklungsaussichten der Saarwirtschaft im deutschen und westeuropäischen Wirtschaftsraum. Forschungsbericht Nr. 4. Institut für empirische Wirtschaftsforschung der Universität des Saarlandes, Saarbrücken
Sievert O, Häring N, Jank HH, Molitor C, Naust H (1991) Zur Standortqualität des Saarlandes. Röhrig, St. Ingbert
United Nations (2006) World economic and social survey 2006. UN Department of Social and Economic Affairs, New York
Wacziarg R, Wallack JS (2004) Liberalization and intersectoral labour movements. J Int Econ 64:411–439

Is Novelty Always a Good Thing? Towards an Evolutionary Welfare Economics

Christian Schubert

Abstract Evolutionary economists are increasingly interested in developing policy implications. As a rule, contributions in this field implicitly assume that policy should focus on the encouragement of learning and innovation. We argue that, from an individualistic perspective, this position is not easy to justify. Novelty and evolutionary change have in fact a rather complex normative dimension. In order to cope with this, the evolutionary approach to policy-making needs to be complemented with an account of welfare the background assumptions of which are compatible with an evolutionary world-view. Standard welfare economics is unsuited to the job, since the orthodox way to conceptualize welfare as the satisfaction of given and rational preferences cannot be applied in a world in which preferences tend to be variable and incoherent. We argue that, in order to deal with the specific normative issues brought up in an evolving economy, welfare should be conceptualized in a procedural way: At the individual level, it should be understood as the capacity and motivation to engage in the ongoing learning of instrumentally effective preferences. Evolutionary-naturalistic insights into the way human agents bring about, value, and respond to novelty-induced change turn out to be a valuable input into this extended concept of welfare. Finally, some implications of this concept are explored.

1 Introduction

Innovation is a two-sided phenomenon: While it is generally beneficial in many senses of this word, it also tends to come with harmful side-effects for some of the individuals affected. It involves "creation" as well as "destruction" in terms of

Reprinted from Journal of Evolutionary Economics 22(3), 585–619, Springer (2012).

C. Schubert (✉)
Max Planck Institute of Economics, Kahlaische Str. 10, 07745 Jena, Germany
e-mail: schubert@econ.mpg.de

G. Buenstorf et al. (eds.), *The Two Sides of Innovation*, Economic Complexity and Evolution, DOI 10.1007/978-3-319-01496-8_11,

increased uncertainty, anxiety, devaluation of human capital, dislocation, status loss, etc. Moreover, we cannot have the former without essential parts of the latter. Hence, rather than being unconditionally desirable, innovation and innovation-driven change have a complex normative dimension.[1] This insight bears an important, but hitherto largely neglected, implication for the development of evolutionary policy advice: We cannot recommend policies to foster learning, change and innovation unless we can make a convincing case that this indeed enhances the actual *well-being* (or welfare) of the agents directly affected. That is why evolutionary economics needs a concept of well-being, which is the first proposition that the present paper is making. The second is that, on the way towards such a concept, evolutionary economists cannot use the well-beaten Paretian path to welfare, since their basic assumptions are at odds with those of the orthodox "New Welfare Economics". Hence, evolutionary economics needs to be creative and to develop its *own* notion—or set of notions—of human welfare. Third, evolutionary economics does have at its disposal the theoretical resources to do this. The present paper outlines some considerations that have to be made when looking for an evolutionary path to a concept of well-being, as well as one possible way to achieve it. To our knowledge, it is the first to do so on a systematic basis.[2]

Most evolutionary economists will be skeptical at first. Thinking about policy goals and welfare (or even "progress"?) in an evolutionary setting seems misplaced—after all, "evolution has no goal"—so why can't we just focus, in a value-free way, on the best ways to foster innovation?

As to the first objection, consider the following. From an evolutionary (in particular, a "Darwinian")[3] perspective it may be tempting to focus on the level of the economic *system* and the way it evolves over time (Sen 1993). On that level we obviously cannot discern any "goals": The system evolves in an open-ended, irreversible and largely unpredictable way. This perspective leads one easily to focus on the system's qualities, such as its capacity to process knowledge, to generate novelty or to adapt to changing environmental conditions. In order to maintain capacities such as these, one may argue that any losses occurring at the level of individual agents just have to be accepted in an unconditional way. In the sufficiently long run, it seems safe to assume that almost everyone will be

[1] This is reflected, throughout economic history, in outbreaks of resistance against economic change. See Witt (1996), Mokyr (2000).

[2] Evolutionary economists have not been *entirely* silent on normative issues. Sartorius (2003: ch. 6) and Binder (2010) discuss notions of welfare from an explicitly evolutionary perspective. Earlier contributions include Dopfer (1976: 19–29), Nelson (1977: 129–143), Nelson (1981), and Nelson and Winter's (1982: ch. 15) largely ignored programmatic call for a "normative evolutionary economics". See also Hodgson (1999: ch. 11), Witt (1996), Witt (2003: 89–91), and the sections on "progress" in Van den Van den Bergh and Kallis (2009).

[3] In the following, we will refrain from using "Darwinian" language in the sense of metaphors or analogies from Darwinian biology. The jury is still out on the issue of whether notions such as "variation" or "selection" can be used in evolutionary economics without biasing the analysis in undesirable ways (Witt 2008; Aldrich et al. 2008). Biases may be particularly strong in the realm of normative reasoning, as this kind of analysis is unrelated to anything known in biology. See also FN 6, below.

compensated somehow[4] (Hick's notion of "general compensation").[5] Laissez-faire is the obvious implication.[6]

It is of course perfectly legitimate to adopt such a system-based view, but we will argue that, being at odds with a basic tenet of evolutionary thinking, it would be hard to defend. Evolutionary economists typically reason in terms of populations of interacting heterogeneous agents. They put emphasis on the agents' idiosyncratic tastes, beliefs and skills, and the way this heterogeneity spurs creativity and learning. In light of these considerations, it seems appropriate to adopt an *individualistic* viewpoint not only in positive theorizing (Witt 1987), but also in matters related to policy advice and normative analysis, i.e., to take seriously the impact of economic change on the quality of life of individuals. This does not mean that systemic considerations are irrelevant. From an evolutionary perspective, the impact that some policy has on the capacity of, say, the price system to process decentralized knowledge matters a lot for the well-being of the individuals affected. The starting point, though, should be to consider changes in well-being at the micro level. This in turn implies that, as soon as there are "destructive" side-effects of evolutionary change that affect some individuals in a negative way (as perceived by the individuals themselves), the institutional framework within which this change takes place is in need of legitimization. It cannot be justified by exclusively referring to some superior qualities of impersonal supra-individual processes *per se*. Rather, the case has to be made that the (given or some alternative) institutional framework is in the best interests of the individual "losers" as well, i.e., that it promotes their well-being, *properly understood*. In that sense, we suggest following Kenneth Boulding's (1981: 195) suggestion that "from the human point of view, progress is improvement in the *state of persons*. Everything else is an intermediate good."[7] This is the basic value judgment of the evolutionary approach to welfare suggested (as a hypothetical desideratum) in this paper.

As to the second objection—"why can't we just focus, in a value-free way, on the best ways to foster innovation?"—two remarks are in order. First, most readers of this journal will be aware that there is, in evolutionary economics, a burgeoning literature concerned with implications related to innovation, technology, IPR,

[4] Strictly speaking, it is a logical fallacy, though, to predict that innovation will continue to benefit the "average" individual in the future. After all, the future tends to be unknown (Witt 2003).

[5] Note, however, that insofar as this concept relates to the criterion of Kaldor-Hicks efficiency, it is highly contentious. See FN 33, below.

[6] As Hodgson (1993: 25) observes, "[i]t is still widely assumed that evolutionary thinking involves the rejection of any kind of state meddling, subsidy or intervention, and the support for laissez-faire on the basis of the idea of 'survival of the fittest'". Notice that Hodgson himself does neither share this assumption (on which see also Whitman 1998) nor endorse any such kind of policy.

[7] My italics. Interestingly, the temptation to oscillate somehow between the systemic and the individualistic perspective can be studied in the cases of both Hayek and Schumpeter. See Vanberg (1994b) and Schubert (2009), respectively.

regional, and entrepreneurship policy.[8] A closer look reveals, though, that this literature is almost exclusively concerned with what John Neville Keynes called the “art of political economy”, namely, the branch of instrumental economics that takes some policy goal as given and examines the most effective means and tools to realize it (Keynes 1917).[9] Here, the role of the economist can be compared to that of an engineer who offers technological advice (although the economist’s task will usually be much more complex). In order to yield consistent and convincing policy advice, though, instrumental analysis cannot refrain from explicit normative reasoning, for essentially three reasons: First, policy interventions typically affect a variety of different policy goals, which implies that trade-offs between these goals may emerge and have to be resolved. Second, over time, policy-makers may learn about new means-ends relationships which will affect the relative plausibility of competing policy goals and criteria, again calling for a normative re-balancing or even a reconceptualization of single concepts. Third, without normative reflection, instrumental theories need to import normative goals and criteria from outside, i.e., from either neoclassical economics or policy practitioners themselves. In the former case, neoclassical baggage may be “smuggled in”, rendering the analysis inconsistent; in the latter case, goals and criteria may be underspecified and in need of further refinement and operationalization (Nelson 1977: 27–29, 148–53). In both cases, explicit normative reasoning is obviously helpful.

There is a second, more general, argument for entering normative territory, though. Evolutionary economists have a lot to offer to enrich the mainstream body of welfare economics itself. Given its increasingly abstract character, the latter’s practical relevance has been questioned for decades, and from very different angles (Sen 1977b; Scanlon 1991; Ng 2003; Atkinson 2009). It is in particular its informational basis (allowing the use of ordinal utility information only) that is in urgent need of reform (Sen 1996, 2009). Recently, happiness research and behavioral economics have tackled the problem. They start from the insight that human well-being (“utility”) is too complex to be simply reduced to “choice”. In particular, due to their context-dependency and inconsistency, an agent’s preferences may systematically depart from her well-being. Then, however, orthodox (revealed preference) approaches to welfare become untenable (Sen 1977a; Broome 1996; Köszegi and Rabin 2008). Things get even more complicated when preferences are assumed to change endogenously, as an inevitable concomitant of innovation; then, the even broader set of preference-based notions of welfare have to be qualified (Sugden 2004; von Weizsäcker 2010).[10] As a consequence, a growing literature is about to develop alternative approaches, substituting non-standard dimensions or

---

[8] See, e.g., Metcalfe (2005), Cantner and Pyka (2001), Dosi et al. (2006), Audretsch et al. (2007), Stam (2008).

[9] In contrast, in Keynes’ framework, *normative* economics proper is about the critical analysis of welfare criteria and policy goals per se. We will adopt this distinction in the following. John Neville Keynes (1852–1949), a British economist, was the father of the much better known John Maynard Keynes.

[10] Classic contributions include von Weizsäcker (1971) and Elster (1982).

"currencies" of well-being such as happiness, capability or "opportunity" for preference satisfaction (Kahneman et al. 1997; Sen 2009; Sugden 2007). Evolutionary theorizing on welfare may contribute to these efforts to reconstruct welfare theory in two ways: First, it may deepen the standard behavioral inquiry into "anomalies" by clarifying their evolutionary origins (see, e.g., Cordes 2008). Second, it may provide insights into the determinants of preference variability as a possible source of time inconsistency (Witt 2001). Essentially, both kinds of contributions benefit from what evolutionary scholars know about the origins of human utility and the way perceptions of utility may change over time.

Why does all this matter? In the context of many practical policy problems, welfare economics' standard theorems turn out to be more or less irrelevant: Consider the issue of how to deal with self-defeating behavioral patterns (involving status races, overborrowing, obesity, ...). Is there a case for policy to interfere in processes of social norm or preference formation, e.g. by means of framing or default engineering (Sunstein and Thaler 2003; Thaler and Sunstein 2008; Lessig 1995; Anand and Gray 2009; Frank 2008)? Does the modern welfare state impact preferences in a problematic way, and, if so, how can we make sense of it (Beaulier and Caplan 2007; Henrekson 2005)? On a more theoretical level, consider the issue of how to assess the well-being of persons whose preferences dynamically adapt to the circumstances in which they find themselves—the famous "sour grapes" problem, which plays an important role in development economics (Elster 1982; Sen 1988). In this sense, the present paper also contributes to the general literature on the empirically informed reconstruction of welfare economics.

We take these to be sufficiently strong reasons for evolutionary economists finally to start investigating the normative implications of their positive insights, by elaborating upon an *Evolutionary Welfare Economics*. Intuitively, an evolutionary perspective should make a difference in this domain. As Richard Nelson and Sidney Winter already pointed out in their "Evolutionary Theory of Economic Change" (1982), "it is apparent that an evolutionary view of what *is* going on in the world of firms and industries strongly influences how one looks at the question of what *should* be going on", leading them to conclude that such a view would require a "rethinking of normative economics" (ibid.: 356).[11] As Kurt Dopfer suggested in his thoughts on a "new paradigm" (Dopfer 1976: 19–29), this "rethinking" should proceed along the lines of substituting empirically informed for axiomatic assumptions at the ground level of welfare economics.

Hence, we argue that, in light of what they know about economic behavior and social phenomena, evolutionary economists should develop their own ideas about how exactly "innovation" can be translated into a viable individualistic criterion of welfare, how it can be justified, formalized, operationalized, and possibly weighed against competing criteria. Moreover, the way these issues are analyzed should be based on background assumptions that are compatible with those guiding positive research in Evolutionary Economics. We suggest that, in a sufficiently elaborate

[11] Italics in the original.

form, an evolutionary account of welfare would (one day) be able to tackle questions such as the following:

– Is innovation-induced structural change desirable *per se*—the more so, the faster it proceeds?
– How should novelty-induced losses and inequalities be measured and how can they be justified?
– How should we think about welfare in an evolving economy in which individual preferences (the usual measuring rod of welfare) change endogenously?

For the purposes of this paper, we wish to boil these questions down to the following one: Consider an evolving economy exhibiting creative destruction, where some innovation generates benefits and costs, as perceived by the agents involved themselves. On the basis of which *concept of welfare* (meaning "which currency of benefits and costs") is it possible to come to at least a conjectural *justification* of the fact that some individuals bear the costs, while others get the benefits? Put much more simply: Why should the individual loser willingly accept his losses in the innovative market game (and, hence, continue to support the game)? Let's call this the "evolutionary compensation puzzle".

The paper is organized as follows. Section 2 shows why instrumental evolutionary economics would benefit from an evolutionary account of well-being. It then discusses the issue of "innovation" as a policy goal. Section 3 argues that, in order to cope with the complex normative dimension of innovation, the evolutionary concept of well-being has to be used within a broader calculus of justice. Since neither the Paretian nor the Hayekian approach to do this calculus can be consistently applied within our setting, we need to resort to the contractarian account that conceptualizes justice as "general consent". Section 4 provides some substance to the evolutionary approach to welfare by presenting a psychologically informed model of preference formation. In light of the normatively relevant implications derived from this model, it then introduces a partly procedural account of well-being as "effective preference learning" and discusses the way it relates to several non-standard accounts of well-being that have recently been discussed in behavioral economics. While the construction of a full-fledged operational criterion of policy-making is beyond the scope of the present paper, Section 5 concludes by briefly presenting some hints as to possible conceptual and practical policy implications.

2 The Evolutionary Perspective

In an evolutionary perspective, capitalism should be understood and appreciated as a restless, self-transformative "engine of growth" (Nelson 1990) rather than a mechanism to generate stable equilibria by having rational agents economize on scarce resources. Accordingly, when evolutionary economists reason about the political and, in particular, the normative (i.e. welfare) implications of their research agenda, there are two issues finding almost general agreement. First, most scholars in the field concur that the key concepts of orthodox welfare economics cannot be used to evaluate phenomena that emerge in an evolving economy. Schumpeter stressed this point repeatedly.[12] The intuition is that the comparative statics of the well-behaved Paretian utility calculus cannot capture the turbulence that characterizes evolutionary "gales of creative destruction".[13] Hence, a formidable research program emerges here. Second, there is widespread agreement that reasoning about policy or welfare implications has to come to terms with the problem of genuine *uncertainty* that characterizes any complex, endogenously evolving economy (Hayek's theme). This immediately rules out pet notions of old-school normative economics, such as the omniscient social planner who directly targets singular allocative outcomes. Hence, the emerging research program appears to be non-trivial.

In what sense are the differences between the neoclassical and the evolutionary approach to economics relevant for this research program? Mainstream economics is, to paraphrase Robbins (1935: 16), about how rational agents economize on a given and closed set of scarce resources. When put this way, the economic problem has a straightforward normative solution: Optimality is reached as soon as the economizing has been done *in extremis*, with all resources being allocated in an efficient and stable equilibrium. According to the first theorem of Welfare Economics, this state of minimal waste is realized under conditions of friction-less perfect market competition. The positive and the normative perspective are closely linked, for the Robbinsian approach to frame the problem of economic life has a latently normative flavor (Witt 2004): Economizing on scarce resources is seen as inherently desirable.

This focus on allocative efficiency rests on a peculiar notion of well-being. In the wake of the ordinalist revolution of the 1930s, the "New Welfare Economics" abandoned the use of cardinal and interpersonally comparable notions of well-being in favor of the strict provision only to accept ordinal utility data as the basis of welfare judgments. "Utility" was stripped of its traditional psychological connotations and came to be defined as the technical degree to which one's preferences are satisfied (Witt 2000; Bruni and Sugden 2007). Accordingly, welfare came to be defined as the satisfaction of given and perfectly consistent (i.e. complete, transitive, reflexive and stable) preferences,[14] with "preferences" in turn being indicated and, ultimately, *defined* by choice. This implies that, by assumption, whatever

[12] See, e.g., Schumpeter (1942: 144), Schumpeter (1954: 130, 1039). In fact, the utilitarian legacy of standard welfare economics has been argued to be incompatible with an evolutionary world-view (Hodgson 1999: ch. 11). See also Hayek (1978: 90–91).

[13] See Ludwig Lachmann's quip that "[i]n a world of unexpected change economic forces generate a redistribution of wealth far more pervasive and ineluctable than anything welfare economists could conceive" (Lachmann 2007: 81, FN 6).

[14] Harsanyi (1982) is the *locus classicus* for this way of thinking about welfare. See Sugden (2009) on the strong (implicit) value judgments inherent in accounts of this sort. Notice that the orthodox account of welfare is explicitly unconcerned with mental states: "satisfaction" does not refer to any

people choose makes them better off.[15] Decades after Samuelson (1938) introduced it, the revealed preference approach still constitutes the hard core of standard welfare economics. With the concomitant restrictions in place, the Robbins-Samuelson framework effectively predetermines a quite narrow set of possible welfare criteria, viz., the criteria of factual and potential ("Kaldor-Hicks") Pareto efficiency (Sen 1996).

What about the evolutionary way to think about economic issues? The literature on the conceptual basis of evolutionary economics (e.g., Nelson 1995; Witt 2008) rejects the Robbinsian conception in favor of a more complex view: Economics is understood as being about the way heterogeneous, boundedly rational and creatively learning agents bring about novelty and change at all levels of economic life. This reflects the commitment to population (as opposed to typological or reductionist) thinking (Andersen 2004). Novelty-induced change tends to be surprising and unsettling. Stable equilibria, should they ever appear, are of little interest, both in a positive and in a normative sense—as Hayek (1948: 101) puts it, "[a]ll economic problems are created by unforeseen changes which require adaptation." Individuals experiment and learn in a time-consuming (and necessarily "wasteful") process which involves their acquiring new preferences, and which, given appropriate institutional conditions, leads to the creation of coordination-based dynamic *order* rather than equilibrium.

Clearly, there is no place for an argument analogous to the mainstream *in extremis* reasoning here, for the background setting is irreducibly *open*. This is a fundamental ontological distinction that, again, relates to the basic difference between the analytical focus on populations as opposed to the neoclassical focus on representative "types". Thus, the evolutionary way to frame the economic problem avoids the strong normative connotations of the Robbinsian approach. In a world of flux, there is no perfectly stable and desirable state of optimality that could conceivably ever be reached. Related to this, there is no fixed position outside the evolutionary process from which to assess its course and results. There is no "given" concept of welfare.

It seems that the only thing that can be said from an evolutionary perspective is that normative issues potentially arise whenever agents confront change. As a consequence of change, their values, goals and aspirations change as well. This is a necessary implication of genuine uncertainty which forces agents to learn; analytically, it can be captured by the notion of endogenous preference change. As a consequence, it does not make sense to conceptualize welfare as the

---

"feeling" of, e.g., "pleasure" (as in classical utilitarianism) but rather denotes a purely technical measure of degree.

[15] Methodologically, this approach was motivated by the supposed lack of reliable scientific knowledge about the process and content of human preference formation. With the rise of evolutionary and behavioral economics (as well as neuroeconomics), this caveat has become obsolete: "non-choice" sources of information on the structure and content of individual preferences are now available and accepted as meaningful by most economists. See Sen (1977a) for the methodological issues involved.

satisfaction of "given" preferences. This is the first step towards an evolutionary account of welfare. It has far-reaching implications: When preferences change—partly as a response to changes in the economy—these preferences cannot be used as a measuring rod for evaluating states and processes of the economy. With the basis of normative assessment in flux, a broad set of standard welfare criteria is called into question, in particular the criteria of Pareto and Kaldor-Hicks efficiency. As yet, it is unclear which alternative criteria could fill the void.

On the other hand, there is much more clarity at the level of *instrumental* research: Given its well-established repository of insights into the determinants of economic behavior and the working properties of institutions, evolutionary economics allows us to derive a rich set of "policy implications" at this level. These tend to have a specifically evolutionary flavor. For, given any pre-assumed policy goal, possible and effective ways to reach that goal depend strongly on whether a neoclassical or an evolutionary model of the economy is assumed to apply in the background (Van den Bergh and Kallis 2009).

Thus, it is hardly surprising that most evolutionary economists, when examining policy-related issues, have tried to remain safely on instrumental ground. There, research typically proceeds in three steps:

First, much ink has been spilled over the fact that, under the difficult epistemic conditions of an evolving economy, policy-making can only be of an experimental ("adaptive") nature, because otherwise it would not be capable of learning (Pelikan 2002; Kerstenetzky 2007). It does not make sense to frame the analysis in terms of some omniscient social planner who "fine-tunes" specific allocative and distributive outcomes. This is so obviously true that no further elaboration is necessary.[16] It has, however, important implications: Standard optimization techniques in normative reasoning have to be eschewed in favor of a framework of comparative institutional analysis. Through observing real-world policy experiments, this analysis allows us to gather valuable instrumental knowledge about the working properties of alternative *rules*. Focusing on rules implies focusing on *patterns* of outcomes that are channeled by these rules. The scientific observer does not address a (fictitious) benevolent social planner. Rather, he shares his knowledge with the citizens themselves (as the principals of policy-making) in a process of public deliberation.[17] Instrumental analysis cannot do without normative orientation, though. If we want to know which institutional arrangements are "better" suited

[16] Nevertheless, this first step of instrumental reasoning is often lacking in applied evolutionary economics research, when (implicitly) an omniscient social planner seems to lurk in the background; see Wegner (2005) for some critical thoughts on this.

[17] See Broome (2008): "Democracy has at least two departments. One department is decision making, and here democracy requires that the people's preferences should prevail... Another department... is the forming of people's preferences...Our preferences about complex matters depend on our beliefs, and democracy requires a process of discussion, debate and education, aimed at informing and improving people's beliefs... The role of economists in a democracy belongs in the second department, not the first... Economists should aim to influence preferences, not take preferences for granted."

to serve the citizens' interests, we need some (however tentative) notion of what constitutes these interests. In an evolutionary setting, this is a non-trivial question.

Accordingly, the second step usually involves the assumption that some policy goal or standard applies. Most often, we find goals to be defined in an eclectic way. They then include the usual suspects, such as "growth", "output", "productivity" or "economic development". Consider Schumpeter's own approach to policy-making,[18] most of the literature on entrepreneurship policy,[19] as well as many scholars engaged in applied Evolutionary Economics who resort to "historical standards". Foss (2006) aptly describes this strategy: "Identify a historical success story ..., and let the qualities that characterize this success story become the relevant standard." Importantly, the policy goals are almost never defended by explicit reference to some underlying notion of welfare (ibid.). Since they seem to do the job anyway, the third step typically involves the comparative analysis of alternative policy tools to achieve the given goal. Hence, at all three levels, instrumental analysis is supposed to refrain from any explicit, even merely hypothetical, endorsement of value judgments.

We argue that evolutionary economists interested in policy advice should not limit themselves to this "pure" instrumental level, for essentially three reasons.

First, policy interventions typically affect a variety of different policy goals. Consider, for instance, entrepreneurship policy. To the extent that it effectively fosters entrepreneurial activity, it can be argued to have an impact on employment generation, employment dynamics, the quantity and quality of innovation, productivity, GDP growth, and the subjective well-being of entrepreneurs and their employees (Van Praag and Versloot 2007). Given this plethora of effects, policy-makers have to decide upon the relative weights to be given to each of them. Moreover, complex trade-offs may emerge and have to be solved. Given their knowledge about the relationships between the different goals, evolutionary economists are well-placed to provide advice in this respect. Clarifying competing goals may even facilitate political agreement.[20] This exercise will certainly be easier for those who are aware of the normative qualities of (i.e. the concepts of welfare underlying) competing policy goals.

Second, an evolutionary approach to policy-making has to account for the indirect side-effects of policy measures (Pelikan 2002). These side effects are sometimes

---

[18] See his suggestion in "Capitalism, Socialism and Democracy" that "we shall call that system relatively more efficient which we see reason to expect would in the long run produce the larger stream of consumers' goods per equal unit of time" (Schumpeter 1942: 190, italics omitted) and the fact that here and in related work he never really elaborates upon this notion of efficiency or welfare. On Schumpeter's complex, insightful and often contradictory reasoning about welfare, see Schubert (2009).

[19] See, e.g. Baumol (1990) who defines as "productive" those kinds of entrepreneurship that stimulate growth and productivity.

[20] See Nelson (1977: 18): "A powerful normative structure can help in the sorting out, weighing, and education of values, and thus can facilitate agreement among groups even in situations where agreement originally seemed implausible."

perverse, often surprising, and always hard to track. This is due to the fact there will virtually always be agents who are negatively affected by a given policy measure. Reacting to policy-induced losses, these agents will try to find creative—and, hence, unpredictable—ways (at least partially) to circumvent these effects (Wegner 1997, Witt 2003: 84–86). Policy-makers have to absorb these surprises and to adjust their instrumental knowledge accordingly (Kerstenetzky 2007). New instrumental knowledge will, however, naturally affect the normative knowledge of the individuals involved, i.e., lead to a reassessment of the plausibility of alternative policy goals and criteria, which again calls for explicit normative reasoning.[21]

Third, the eclectic approach to policy-making described above runs the risks of adopting, without critical (normative) reflection, policy goals and criteria that are underspecified ("output"), contentious in their own right ("GDP growth")[22] or reflect non-evolutionary background assumptions ("efficiency", "market failure"). The last case is particularly troubling. Consider the notion of "market failure", which is often used in order to justify entrepreneurship policy. As a standard, it is derived from the concept of a "perfect" end-state. Such a state, though, lacks meaning in an evolutionary context (Metcalfe 2001). What looks like market imperfections from a neoclassical perspective (such as, e.g., asymmetric information), is a perfectly natural quality of evolutionary processes. In an evolutionary setting, market "failures" provide the incentives necessary to engage in explorative behavior and entrepreneurial activity. Lack of normative reasoning may lead theorists to borrow concepts from standard welfare economics, which in turn risks rendering the instrumental theory inconsistent.[23]

Thus, in an evolutionary world, there is no pure instrumental "engineering". Normative reasoning seems to be required in order to ensure the consistency (and increase the relevance) of evolutionary economics' policy advice, and it should be done explicitly rather than implicitly. The first step would be to sketch the contours of a possible dynamic notion of well-being that is compatible with basic

[21] See Witt (2003: 87): "The policy maker gathers information to learn about the consequences [of some policy]... For example, if the proclaimed goal is a 'more just' income distribution, it is only with the experience made with some concrete redistributive policy measure that the policy maker finds out what kind of 'justice' actually results (which is a factual question) and whether its observed consequences are indeed considered to be worthwhile (which is a normative question). It may turn out that other goals are also being affected... The experience made will most probably be one of trade-offs which may induce ... revaluations at the level of the goals. A prominent case are inconsistencies... between the attainment of short run goals and the attainment of long run goals which are discovered only later." Note that all this does not affect the categorical (logical) distinction between positive and normative statements.

[22] "GDP growth" as a policy goal is subject to intense critical scrutiny, particularly in the field of ecological economics.

[23] This third argument relates to the post-positivist insight that there really is no perfectly "value free" way to engage in instrumental analysis in economics (Myrdal 1933). The scientist's choice of goals inevitably directs the attention of the policy-maker or citizen who is addressed (Witt 2003).

evolutionary assumptions. The starting point would arguably be to turn “innovation” or “learning” into a viable concept of welfare.

Let’s consider what can be found in the literature. In his vision of “evotopia”, Hodgson defends the view that policy should aim at establishing conditions that foster learning and encourage innovative behavior at all levels of the economy (Hodgson 1999: ch. 11). In the words of Metcalfe (2001: 561–62), the economic process should be evaluated according to how far it succeeds in “overcoming ignorance”. This, he argues, is a standard that would be more “exacting” than the conventional criteria of welfare economics. Potts (2001: 428) briefly pleads in favor of enlarging people’s access to knowledge. Views such as these seem to reflect an older idea developed by Douglass North (1990) in the context of his analysis of institutional arrangements: In a dynamic economy, he argues, social states should be judged according to their “adaptive efficiency”, a criterion which he defines, somewhat vaguely, as being “concerned with the willingness of a society to acquire knowledge and learning, to induce innovation, to undertake risk and creative activity of all sorts, as well as to resolve problems and bottlenecks of the society through time” (ibid.: 80). Although the concept has been much cited ever since, it has never really been systematically elaborated upon as an element of an evolutionary approach to normative economic theorizing.

Thus, it seems that evolutionary economists, when moving beyond the purely instrumental level of analysis and beyond arbitrary normative eclecticism, tend to endorse “innovation” as a criterion of “welfare”. Hence, any policy that promotes innovation is to be welcomed. We will take this to be the starting point of our inquiry. It cannot be more than a starting point, though, for the following two reasons:

(i) As already noted in ‘Section 1’, “innovation” may provide a valid notion of welfare from a purely system-based view (Sen 1993), but not from an individualistic perspective. To illustrate the tension between these two views, consider Hayek’s positions on the policy implications of an evolutionary world-view (Kerstenetzky 2007; Vanberg 1994a; Sugden 1993).[24] On the one hand, Hayek typically argues that the common good should ultimately be conceived as the

[24] On a more applied level, Potts (2004) stresses the importance of the systemic view on welfare in the context of assessing financial “bubbles”. Starting from the premise that bubbles are “natural mechanisms of economic growth and evolution”, he argues that they are “ultimately a sign of system health and vigour, not of decadence and decay” (ibid.: 16), and that policy should appreciate their beneficial effects: “Inside a bubble, the cost of experimentation, and therefore variety generation, is lowered and,... the process of structural change is accelerated... Learning is accelerated within a bubble, and radically new business ideas can get a start, as can radically new products” (ibid.: 18–19). His policy implications are straightforward: “Policy should not worry about bubbles; if anything, and where it is safe to do so, it should perhaps even encourage them.” (ibid.: 20). In light of recent events in global financial markets, this may sound absurd. Potts, however, has a point in stressing the systemic aspect of welfare: In order to establish the conditions necessary for an evolving economy to work smoothly, bubbles do have some functional properties. The argument gets into trouble, though, as soon as the normative and the instrumental levels are mixed: Fast learning and variety generation would then be judged as desirable goals *per se*. Potts

"facilitation of the pursuit of unknown individual purposes" (Hayek 1976: 1).[25] In his later writings, however, he tends to switch to supra-individual criteria (such as "population size"), arguing that given the way impersonal processes of socio-economic evolution work, it would be meaningless to ask whether the individuals affected by them would ever agree to the results, since "in any case, our desires and wishes are largely irrelevant" (Hayek 1988: 134).[26] Positions such as these can, of course, be read as either implying "evolutionary agnosticism" (Vanberg 1994b) or a notion of "welfare" that is detached from the level of individual benefits. In the latter case, the focus would be on maintaining certain *systemic* qualities of the innovative market game. Whether the individuals affected like its impact on their personal well-being or not, these features have to be protected against any attempt to push centralization too far. In contrast, from an individualistic perspective, innovation certainly has an *instrumental* value in enhancing the individuals' well-being. In order to establish this case, though, we have to clarify in which sense exactly the well-being of immediate *losers* is affected.

(ii) Even apart from any "destructive" side-effects, the idea to "maximize innovation" turns out to be a self-defeating concept. It may be possible to provide an institutional environment that promotes innovative or entrepreneurial activity to the greatest extent possible. Note, however, that an unconditional acceleration of innovative change may turn out to disrupt the ordered relationships that are necessary for explorative activity in the first place. Too rapidly changing environments will, in the end, stifle human creativity (Hodgson 1999: 248). Moreover, other important policy goals may be jeopardized: The sciences tell us that, while stimulation through novelty is an important source of human well-being, there is an upper bound: "extremely novel situations... are experienced as unpleasant and trigger avoidance behavior" (Cordes 2008: 129). Ultimately, too rapidly changing environments may produce discomfort (Bianchi 2002: 12) and undermine the perceived legitimacy and societal acceptance of innovation-friendly institutions.

From this, two preliminary conclusions follow. First, in order to translate into a viable concept of welfare, the concept of "innovation" has to be embedded, as it were, into a broader set of considerations related to the impact of economic novelty on human well-being. In this form, a "complete" notion of innovation-induced welfare will then be able to cope with the evolutionary compensation puzzle referred to above: How can we reconcile an institutional and policy framework

---

seems to be aware of this risk of producing counter-intuitive policy advice (note his caveat: "*where it is safe to do so*"), but does not elaborate upon this issue.

[25] See also the references given by Vanberg (1994b: 465–66).

[26] Hayek (2009) constructs a strong moral obligation for man to support "rapid material progress", which leads him ultimately to argue that "we are ... the captives of progress; even if we wished to, we could not sit back and enjoy at leisure what we have achieved" (ibid.: 52). On this tension in Hayek's normative argument, see also Vanberg (1994a, b: 183) and Gray (1999: 154).

that *grosso modo* encourages innovation with the individuals' desire to avoid excessive hardship? Solving this puzzle is the key to (i) the pragmatic problem of absorbing resistance against innovation-friendly policies, and to (ii) the normative problem of providing at least a tentative justification for these kinds of policies.

Second, given that the "broader set of considerations" referred to above build upon the notion of innovation and the concept of an evolving economy, it seems obvious that individual well-being can only be understood and modeled in a (at least partly)[27] *procedural* way. This means, on the one hand, following an intuition expressed by Hayek in his "Constitution of Liberty", that "progress" should not be conceived in static terms, but rather dynamically, as "movement for movement's sake" (Hayek 2009: 38): Translated in preference terms, the successful learning and refinement of ever new preferences might be used as the key element constituting individual well-being, independent of specific degrees of preference satisfaction that result from the process of learning. On the other hand, the acquisition of new preferences in itself does not qualify as a sufficient indicator of an increase in well-being, as this would again boil down to a view that anything new is desirable. Rather, we have to probe, *within the preference learning framework*, whether individuals are negatively affected by economic change in a *systematic* way. As we will see in Sections 4.2 and 5, below, this may be achieved by focusing on the possibility that a dysfunctional institutional market framework may systematically deprive a subset of individuals of the ability to engage in "effective preference learning" or even undermine the sustainability of individuals' processes of preference learning altogether.

3 A Problem of Justice

Having outlined the broad contours of a notion of well-being that would fit into an evolutionary framework, the next step consists in linking this notion to the "evolutionary compensation puzzle" which is at the core of the normative issues brought up by an innovation-driven economy. It is important to keep in mind that the basic purpose of an evolutionary account of well-being is to cope with this puzzle. Joseph Schumpeter, while typically shying away from any systematic elaboration of the normative dimension of the "process of creative destruction", had a clear intuitive grasp of the complexities involved, as evidenced by this particularly colorful statement:

> "[A] process of degeneration, of degradation of large circles (of society) accompanies the upward movement... Large circles see their economic basis being pulled away. This does

[27] *Purely* procedural criteria would neglect any considerations related to outcomes (or patterns of outcomes) whatsoever, thereby running the risk of generating strong counter-intuitive conclusions. They are hardly ever consistently defended (Nozick 1974 being a notable exception), which is why we will not consider them here.

> not happen abruptly, but gradually. Through generations, the people affected live a deprived existence full of hopelessness. Their moral and intellectual powers dwindle, the more so the more the economic atmosphere they find themselves in is darkening.
>
> An observer from outer space wouldn't notice these phenomena, so fascinating is the development at large—and those losses are just their reverse. They are due to the fact that the services these agents offered are now being offered in a better way. Even the suffering thus caused serves to get rid of the obsolete and to impel novelty. Those who are playing the drama, however, as well as those observers close to them, think differently about it. They cannot ignore the shouting of the crunched who are crushed down by the wheels of novelty."
>
> (Schumpeter 1912: 503)[28]

Apart from its unquestionable beneficial impact on the "average" individual's welfare (however broadly defined), human innovativeness typically also involves insecurity, risk and potentially severe detrimental consequences which in turn may fuel resistance against novelty (Witt 1996, 2003).[29] The costs incurred may be reflected in changes in the relative price vector ("pecuniary externalities") or not ("technological externalities"). Winners may outmatch losers, losses may be "short-term", costs and benefits may accrue to the same or to different people over time—everything is possible. In the sufficiently long run (where "long run" may very well refer to decades), it is almost trivially true that everyone benefits somehow—at least if they are alive to enjoy it. Notice, though, that simply referring to some unspecific "future benefits" is hardly sufficient to solve the justificatory issues involved here. Rather, an evolutionary theory of welfare would have to give a well-justified answer to the question of how the benefits and costs associated with innovative change should be conceptualized, weighed, and balanced, whether and how losers should be compensated, and under what conditions gains through innovation are worth pursuing.

First of all, we have to examine the ambivalent nature of evolutionary economic change in more detail. A closer look at the underlying process of social learning reveals that simply equating welfare with innovation-induced "growth" misses something that is not only essential as a positive explanandum, but also in terms of its normative relevance. Evolutionary processes are based on the generation and diffusion of knowledge. Progress in knowledge, however, is "necessarily *non-uniform*" (Metcalfe 2001: 565).[30] For economically relevant knowledge is always prone to falsification and obsolescence when circumstances change. In an economy that operates outside a state of equilibrium, there are at all times "internally generated reasons for beliefs to change" (ibid.). With the epistemic basis of economic behavior in permanent turmoil, though, it follows that economic change itself necessarily proceeds in an uneven way: As Metcalfe (ibid.: 564) puts it, "advance in some directions is associated with deterioration in others", even

---

[28] My translation from the German original.

[29] Cf. Buchanan's(1977: 27–30) related hint at the possibility of "spontaneous *disorder*".

[30] Italics added.

inevitably so: "Creative destruction implies the destruction of some activities as a necessary element in the growth of others" (ibid.: 566).

Given the necessarily uneven nature of knowledge-based economic evolution, it appears now that what is missing in the evolutionary account of policy-making is the recognition that, when observing and studying processes of "creative destruction", we are facing a problem of how to deal with *imbalances*. This is an intricate issue: On the one hand, imbalances are a naturally occurring phenomenon in any evolutionary system. Consider Schumpeter's entrepreneur, whose activities, by introducing novelty into the system, necessarily have a disequilibrating impact on the economy.[31] Old production processes disappear, incumbent firms find themselves displaced. The imbalances implied here hardly constitute a normative problem *per se*. On the other hand, ultimately some of these imbalances translate into imbalances on the level of individual scores of well-being. From an individualistic perspective on welfare, it is at this point that they become normatively relevant by indicating what amounts to a problem of *justice*.

We cannot do without a concept of justice when questions about trade-offs between novelty-induced gains and losses—in the (yet to be determined) currency of well-being, which now translates into a currency of justice—have to be solved. This holds independently of whether gains and losses accrue within one time period or are spread over several periods: Typically, but not necessarily, novelty-induced "destruction" is imminent, while gains are realized in subsequent periods. Some concept of justice underlies (often implicitly) any logically possible answer to the evolutionary compensation puzzle.

Consider the "classic" answers that have been developed in normative economics: Actual Pareto efficiency (where everyone has to be fully compensated for any losses) represents one polar position, while the opposite position would be defined by a libertarian view according to which *any* welfare risks involved in an innovative market setting have to be tolerated.[32] Somewhere between these extreme positions we find the standard of *potential* Pareto (or "Kaldor Hicks") efficiency, defining any allocative change as welfare-improving that leaves the winners with the mere possibility of fully compensating the losers.

Can these criteria be used to cope with the evolutionary compensation puzzle? The answer is negative, for essentially two reasons. First and most obviously, all criteria generate implausible implications at the practical level of policy-making: Consider actual Pareto efficiency. Compensating any losses whatsoever would immediately stifle any innovative activity. This can hardly be argued to be in the best interests of the losers of innovative change (Witt 1996). As to the opposite extreme position, dismissing any compensation is incompatible with an individualistic perspective, as discussed above. In light of these considerations, Kaldor-

[31] Note that this does not apply to Kirzner's entrepreneur, whose arbitraging brings the economy back into equilibrium.

[32] This holds as long as an egalitarian and non-discriminatory allocation of basic *rights* is maintained.

Hicks efficiency may look like an attractive middle-range candidate. Due to its total neglect of distributive concerns, though, it may also be argued to be incompatible with an individualist position that does not dismiss losses at the individual level *a priori*.[33]

Second, and more importantly in our context, all classic criteria frame the compensation problem in terms of *utility*, to be understood as degrees of preference-satisfaction. Preferences are assumed to be given and perfectly consistent. In an evolutionary setting populated by real-world agents, this way to define the "currency of justice" would obviously have to be modified.[34] Hence, the standard answers to the compensation puzzle cannot be applied in our evolutionary setting.

What is required, then, is a distinct, non-standard concept of justice that tells us under what conditions the losses involved in evolutionary economic development can be seen as "just" and legitimate.[35] It may be legitimate to encourage innovation that generates serious hardship for a subset of the affected population, but it is not *necessarily* so. Rather, this evaluation depends on a normative analysis involving a justice "calculus" that in turn critically hinges on a concept of well-being (as the currency of justice).

To be sure, this concept of justice and its related criteria have to be used within the context of an endogenously evolving economy with its harsh epistemic conditions. This is difficult, but not outright impossible. Specifically, we can use two instruments to achieve this aim. First, a balancing of benefits and costs does not necessarily involve a full-fledged calculus in a manner analogous to a societal balance sheet, where gains and losses are offset across individuals. There are severe epistemological problems (in particular in our evolutionary setting) that make such an approach a non-starter. Consider, for example, the complex question of how to do interpersonal comparisons of well-being levels. It is possible to suggest a solution to the compensation puzzle in the following way. Given an appropriate "currency" of well-being, we can stipulate it to be the sole equalisandum, i.e., that *all* individuals affected by innovative change have to be endowed with a "basic" amount of it. Once this condition is satisfied, everything else can be left open to the vagaries of the evolving economy. Here, the specific notion of well-being enjoys a lexicographic priority, since, in determining how well-off an individual is, it cannot be traded against any alternative currencies of well-being.[36]

---

[33] To paraphrase Rawls' well-known critique, the offsetting involved in applying this criterion does not take seriously "the distinction between persons" (Rawls 1971: 27). Add to this the manifold technical problems widely discussed in the literature (e.g. Scitovsky 1941; Little 1957; Gowdy 2004). Anticipating most of this, Schumpeter (1954: 1072, FN 9) dismisses the Kaldor-Hicks criterion in just one sentence.

[34] To illustrate, consider Rawls' difference principle ("inequality of winners and losers is *just* as long as losers are better off in the regime allowing the inequality than in a regime disallowing it") which frames issues of justice in terms of all-purpose "primary social goods" rather than in terms of preference-based utility (Rawls 1971: ch. 2).

[35] "Just" and "legitimate" are used synonymously in the following.

[36] Again, Rawls' notion of "primary social goods" may illustrate this approach.

The second instrument can be described as follows: The genuine uncertainty inherent in any novelty-generating evolving economy can be seen as a real-world setting close to the hypothetical situation individuals find themselves in behind a "veil of ignorance" (Rawls 1971).[37] This is a key device of the contractarian toolbox, which can be used to guide and to structure the theorist's normative reasoning as well as the individuals' public deliberation about alternative concepts of justice (Buchanan 1977; Vanberg 1994a, 2006). Within such a contractarian framework, we will argue that a concept of justice serving our purposes can in fact be construed under the following two conditions. First, given the genuine uncertainty prevailing in an evolving economy, it should be tailored to the normative analysis of (constitutional) *rules* that generate observable *patterns* of outcomes rather than specific allocative and distributive social states. The underlying intuition is the following: Any normative perspective that starts from an evolutionary world-view has to be at least partly process-oriented in the sense that, "rather than seeking to control outcomes directly, it seeks to affect outcomes *indirectly*, by subjecting the processes from which outcomes emerge to general rules that promise to generate overall desirable patterns of outcomes" (Vanberg 2006: 204).[38] Policy, then, is to be understood as *rule design*. By designing rules, it can influence the direction and intensity of innovative behavior (at least within limits).

Second, the concept of justice should be embedded in an overarching principle and criterion of "general consensus", stating that those rules are "just" that are generally agreed upon by the individuals affected by them. Analogously, rules can be conjectured to be "just" in this sense insofar as they can plausibly be reconstructed as being agreeable by all individuals affected, when these are placed in a situation (such as the veil of ignorance) that reflects widely held intuitive notions of impartiality.[39] Here, "general consensus" would replace the orthodox idea of "social welfare maximization". The latter is fundamentally at odds with the evolutionary world-view, because it presupposes a closed system of parameters that readily provides some maximandum.

Among all conceivable normative criteria, "general consensus" is the one requiring the weakest prior value judgments: It is derived from the basic principle of individualism, according to which it is only the current real-world preferences (and consent) of heterogeneous individuals that carry genuine legitimizing force (Vanberg 2006). Thus, in what follows, the attribute "just" will be understood in the light of a general consensus criterion, meaning that some balance of gains and losses is just insofar as it can plausibly (with good reasons) be taken to be generally acceptable.

[37] See Hanusch and Pyka (2007: 284).

[38] Italics partly omitted.

[39] Cf. Buchanan(1977: 139–40) on the (non-justificatory) status of statements based on such a "hypothetical contract" analysis. Ostrom (2005) illustrates the practical relevance of this way to achieve mutually beneficial systems of rules to solve social dilemma issues.

Under the conditions of an evolving economy with changing preferences, though, the general consensus criterion cannot be made operational without at least a minimum specification of the processes driving the formation of individual preferences in the post-constitutional "game of life". Behind the veil of ignorance, people should be assumed to be aware of the fact that they are going to play an innovative (novelty-generating) market game, in which their own preferences will change continuously over time. Given this, we face the question in terms of the notion of welfare these agents will plausibly deliberate over rules. It is at this point where the evolutionary theory of human well-being steps in. This issue will be examined in more detail in the following section.

4 Preferences and Welfare

Summing up what has been said so far, we suggest classifying the results and side effects of evolutionary change as "legitimate" insofar as they emerge under rules that, under the conditions of the "veil of ignorance", can plausibly be judged acceptable by all individuals affected. In public deliberation, arguments in favor of the legitimacy of rules (in particular rules specifying compensation) have to be framed in terms of whether these rules can plausibly be seen as conforming to the individuals' common "constitutional" interests.

This way to approach the evolutionary compensation puzzle has two important implications: First, we cannot stipulate *a priori* that the detrimental effects of "creative destruction" are somehow automatically or naturally compensated for by the beneficial impact—however conceptualized—of economic novelty and learning. Rather, we have to examine what kinds of losses and gains are involved and how they relate to the well-being of the individuals affected, in order to be able to make any meaningful statement about whether the resulting balance can be qualified as "just" in the sense it is understood here. Second, neither can we state, without further inquiry, that, as soon as the individuals affected have agreed to play a novelty-generating market game, they now have to tolerate *any* costs and risks implied by this (Vanberg 2006: 207). Such a statement would simply beg the question.

In order to make progress, two questions would have to be clarified: First, we would need to specify what is meant by "losses and gains", i.e., which "welfare currency" or "currencies" should be used to express these variables. This is a normative task. Second, given the answer to the first question, the value of the variables would have to be examined empirically. This is a task that goes beyond the scope of the present paper. Hence, in the remainder of the paper, we will focus on the first—normative—issue: How do we conceptualize welfare in an evolving economy?

We have to step beyond the notion of "given" and "rational" preferences. When confronted with a never-ending, highly complex stream of new ideas, goods and services, cognitively constrained individuals cannot be assumed to be capable of forming rational preference orderings at any time. As research in behavioral economics shows, individual preferences tend to be inconsistent, the more so, the more complex the individual's choice environment (Rabin 2002). Moreover, individuals cannot be assumed to possess, at any time in the process, a "complete" set of preferences (Ariely et al. 2006). It makes more sense to assume that they start from an incomplete set and "construct" or, rather, learn new preferences when necessary—this being one way to cope with complexity (Earl and Potts 2004). Individuals adapt to changing circumstances by trying out and acquiring new preferences.

4.1 The Process of Preference Formation

Since these departures from standard preference theory will be essential for our argument, we have to take a closer look at the mechanisms involved. We have to open the black box of preference formation and examine the underlying factors that drive this process. According to the most advanced psychological ("naturalistic") studies on this subject, endogenous preference formation is driven by the continuous learning of new *needs* (Witt 2001). As (non-extensional)[40] behavioral dispositions, needs can be seen as the prime underlying factors shaping the formation of (extensional) preferences.[41]

Needs are the key factor motivating human behavior. Most of them are acquired during the individual's socialization process and life-long learning history. This process of acquisition in turn is based on two genetically anchored factors: Individuals are universally endowed with (i) a relatively small set of genetically hardwired *basic needs* (including the needs for water and caloric intake, but also psychological needs such as the need for social recognition and status),[42] and (ii) a small set of elementary learning mechanisms, such as operant conditioning, which determine the way the individual acquires new needs.

If a basic need is deprived, the human organism reacts by choosing some activity to reduce the deprivation. If successful, the temporary satiation of the need triggers a rewarding (pleasurable) experience. For most basic needs that are relevant for economic behavior, the logic of conditioned reinforcement applies: The rewarding

[40] I.e., not necessarily involving comparisons between alternative social states (such as bundles of goods); see Kahneman and Sugden (2005: 164) for the terminology.

[41] The following five paragraphs rely heavily on Witt (2001) whose account of want learning is inspired by Benthamite hedonism (Kahneman et al. 1997), need-theoretic approaches (e.g. - Georgescu-Roegen 1954; Ironmonger 1972) – without, however, endorsing a 'Maslovian' hierarchy of needs – and the theory of instrumental conditioning (Herrnstein 1990). For methodological details, see Witt and Schubert (2010).

[42] Such needs (or "basic" needs) include the needs for food, water, sleep, warmth, sex, etc., but also psychological needs, such as the need for cognitive arousal and the need to comply with social norms. See Witt (2000, 2001) and Witt and Schubert (2010) for details.

experience increases the rate at which the activity is chosen in the future. Importantly, while any activity that contributes to the satisfaction of such a need causes a rewarding experience, satiation dynamics differ strongly between needs. Some of them, such as those related to the psychological need for social recognition and status, may be almost impossible to satiate fully. Put differently, the individual investment of resources into satisfying such needs yields only limited quantities of rewarding experiences, as compared to the satisfaction of alternative needs.

For our purposes, two properties of the learning dynamics involved are particularly important. First, the learning proceeds partly at an unconscious ("non-cognitive") level. While basic needs originate from the universally shared genetic endowment of man (think about the intake of air, liquids and food), acquired needs arise mainly through a process that involves associative learning. Individuals learn in an associative way when an originally neutral stimulus (say, some alcoholic drink) is repeatedly performed simultaneously with an activity that satisfies a pre-existing need (say, the company of friends). Over time, an association will then be formed between the neutral stimulus and the rewarding experience from satisfying the original need. The formerly neutral stimulus becomes a "secondary" reinforcer, and the individual learns a new ("acquired") need for the originally neutral activity and the consumer items related to it.[43] While this kind of associative learning often happens in an unconscious way, individuals of course also acquire new needs by using their specifically human cognitive capacities. Individuals learn about the characteristics of goods as well as about the goods' ability to satisfy certain needs. This positive consumer knowledge is obtained by personal experiences as well as by imitating and communicating with peers and other consumers and from the use of media and public information services.

This leads us to the *second* characteristic that should be kept in mind: The way the individual acquires new needs also has an impact on the kind of instrumental knowledge he absorbs (about product characteristics, say). Agenda-setting effects play an important role here, since new consumer knowledge is more likely to be perceived and processed the closer it relates to basic or acquired needs that already exist or that have already been learnt, respectively. There is a dynamic interplay between associative learning, on the one hand, and cognitive learning, on the other hand (Witt 2001: 30). The learning of new needs affects the dynamics of knowledge acquisition in society. In turn, an individual's new factual knowledge affects his preferences, leading to the hypothesis that knowledge-induced possibilities of action and preferences co-evolve. As a function of their state of knowledge, people face different states of preferences (different "utility functions") at different points in time (Witt 2003).

Through the processes that have only been sketched here, in the course of a lifelong history of individual learning, a complex hierarchy of ever more refined and sophisticated, often even highly idiosyncratic wants and preferences emerges. By

[43] In the long run, though, the intensity of the acquired want tends to weaken unless the original association with the underlying need is occasionally re-established (Witt 2001: 28–29).

trying out new ways to satisfy their needs and wants, individuals "specialize" in particular kinds of consumer knowledge, preferences, and consumption activities.

4.2 Preference Learning and Welfare

What does all this imply in terms of welfare? The empirically informed account of preference formation that has been described above allows us (i) to draw some general, normatively relevant insights about preference change, (ii) to describe, as precisely as this is possible at this early stage of the analysis, a procedural criterion of preference learning, and (iii) to explore the relationship of this criterion to alternative non-standard concepts of well-being that have so far been suggested in the literature (see 'Section 1', above).

The model described in Section 4.1 tells us, first, that preferences change in a way that is neither perfectly determined nor completely arbitrary. Rather, "laws of motion" can be discerned to a certain degree, which makes it possible to theorize about preference change and to build a normative theory upon the insights gained thereby. Second, preferences here emerge as *instrumental* entities, rather than goals *per se*. As the whole process is driven by basic and acquired needs, it is in order to satisfy *them* that individuals acquire and try out new preferences. Importantly, this reconceptualization of preferences leaves room for the possibility (and, indeed, high probability) that individuals sometimes form preferences that are dysfunctional.[44] For instance, certain needs (such as the one for social recognition and status) are hard to satiate, thereby generating only transient gains in satisfaction. It may even be possible that interdependent preference learning exacerbates this tendency in such a way that the ongoing process of successful learning is effectively undermined.[45] Third, preference learning is closely intertwined with the acquisition of new knowledge on means-ends relationships. Hence, policy may use the provision of knowledge as a lever to gain (limited) influence over the way individuals learn new preferences.

In light of these considerations, we suggest to follow up on the first sketch of a procedural notion of well-being given at the end of Section 2, above. In an evolutionary perspective, well-being may be defined as the individual's *motivation and ability to engage in the ongoing, instrumentally effective learning of new preferences in all domains of life*. The model described in the foregoing sub-section supports the intuition that this process of learning may "work" more or less well, and that the way it works depends in part on the agent's "learning environment" (e.g. the quality of instrumental knowledge to which he has access). For the sake of brevity, let's call this concept "effective preference learning".

[44] See the related evidence given by, e.g., Frey and Sutzer (2007) on the human tendency to mispredict the hedonic impact of consumer decisions.

[45] See Cordes and Schubert (2010) for a formal model of cultural learning that captures this idea.

Three remarks are in order. First, this notion of well-being is tailored to a focus on *patterns* of outcomes rather than particular outcomes. In other words, it is independent of the agent's specific success, in singular circumstances, to achieve high levels of need and preference satisfaction. The continuing failure to achieve even moderate levels of need and preference satisfaction may, however, *indicate* a low level of well-being, as defined in our sense. *In that specific sense*, the individual would be seen as being unjustly affected by negative side-effects of innovative change. He would then have a legitimate claim to being compensated in terms of our currency of well-being. Second, this notion of well-being should be understood as being lexicographically prior to alternative notions. Losses in its dimension (or currency) cannot be offset against gains in other dimensions of well-being. This implies that simply losing income or wealth or human capital in the "gales of creative destruction" does not, in itself, indicate any injustice. It does so only when these losses finally translate into losses in the currency of our concept of well-being. By combining a focus on the paths of preference acquisition with a responsiveness to negative patterns of outcomes, this criterion of well-being is of a procedural nature, but not exclusively so. The way alternative paths are assessed is partly a function of the patterns of outcomes they bring about. Third, this notion of well-being restricts the kind of practical instruments that policy (understood as rule design, see above) may legitimately apply. This concerns the legitimacy of "paternalism", in particular the use of policy tools suggested by the "libertarian paternalist" movement in behavioral economics (Thaler and Sunstein 2008; Camerer et al. 2003). As the space of future preference paths is open, it is up to the individual—e.g. in his role as experimental consumer—to try out new preferences, to discover their instrumental effectiveness, and ultimately to explore new sources of utility. While policy has to make sure that this process "works" reasonably well (by maintaining appropriate institutional preconditions), it is generally[46] ill-placed to do so by restricting the specific *content* of preference learning. Hence, when channeling the impact of innovative change on processes of preference formation, policy should refrain from using instruments that constrain people's opportunity sets.

This way to think about well-being offers three advantages that are immediately evident (two additional advantages will be discussed presently). First, it offers a genuinely individualistic concept by focusing on the preconditions of leading a "good life" and exercising consumer sovereignty at the micro level. It is safe to assume that individuals care about their personal motivation and ability to engage in the ongoing learning of new preferences, as this can be argued to be constitutive of their personal "identity", seen in a dynamic perspective. In terms of the contractarian thought experiment introduced in Section 3 above, one may advance the following proposition: Assume that, when deliberating about rules of the (future) innovative market game from behind the veil of ignorance, individuals

[46] There may be exceptions, e.g. with respect to the consumption of addictive substances (see below).

are aware of the fact that their future preferences will develop according to the theoretical account given in Section 4.1. At the same time, they of course do not know anything about the precise content of their future preferences. Then there are good reasons to hypothesize that they will wish to design the rules in such a way that the conditions for the ongoing effective preference learning are maintained over time. Put differently, there are good reasons to assume that they will think about the "evolutionary compensation puzzle" in terms of our procedural notion of well-being. Related to this, our notion of well-being can be shown to lead to normative arguments and policy implications that are intuitively appealing to evolutionary economists (see Section 5, below). Second, it can be consistently applied in an evolutionary setting, due to its capacity to cope, in a constructive way, with the phenomenon of endogenous preference change (by not treating it as a source of "irrational" inconsistency). Third, its further elaboration and operationalization can benefit directly from both evolutionary-naturalistic insights into processes of learning (as demonstrated above), and from insights offered by non-standard accounts of well-being that have recently been developed in behavioral economics.

In behavioral economics, eschewing the orthodox choice-based concept of well-being has spurred normative reasoning in at least three different directions. We will briefly show how a further refinement of our concept of well-being may benefit from this ongoing research. First, many scholars have rediscovered the classic utilitarian vision of human welfare by elaborating upon a notion of well-being as happiness (Kahneman et al. 1997; Frey and Stutzer 2002). It is suggested to step "beyond preferences" (Ng 2003) in assessing the "deeper" sources of human welfare. Research on happiness or "subjective well-being" (SWB) proceeds mostly in terms of outcomes—asking, e.g., which impact factors such as unemployment, inflation or environmental pollution have on an individual's level of SWB—leading to policy implications that are framed in terms of the static maximization paradigm. For instance, GDP is replaced with some indicator of "gross national happiness" (Blanchflower and Oswald 2005). There is, however, also ongoing research into *procedural* sources of happiness ("procedural utility", Frey et al. 2004). Institutional environments may increase people's happiness by satisfying their need for self-determination. They do so by allowing individuals actively to pursue their self-chosen goals in their own fashion. This idea is obviously related to our focus on the individual's motivation and ability to engage in effective preference learning. Hence, policy-related insights from this strand of the happiness literature may be a source of inspiration for an evolutionary account of welfare.

Second, our notion of well-being bears resemblance to parts of the capability approach advocated by Sen (1979, 1988, 1996, 2009), among others. There, an agent's well-being is defined as a function of what he is able to achieve. Put differently, it is constituted by the vector of functionings that are effectively available to her (Sen 1985; Clark 2006). A functioning is defined, very broadly, as "an achievement of a person: what she or he manages to do or be" (Sen 1980: 10). Examples include "being adequately nourished", "having a basic education", or "being able to appear in public without shame". Functionings refer to the use a person makes of the commodities he commands and his ability to transform commodities into personal quality of life (Clark 2006). An agent's capabilities are defined as the set of alternative combinations of functionings that he is able to achieve (e.g., the ability to achieve the state of "being adequately nourished"). While these elements display a certain outcome-orientation, Sen also refers to procedural aspects that are more pertinent to our evolutionary approach, such as the "substantive freedoms [the agent] enjoys to lead the kind of life he or she has reason to value" (Sen 1999: 87). Hence, an agent's well-being crucially depends on the extent to which he has the freedom to choose among alternative options.[47] In Sen's terminology, our notion of well-being might be characterized as a "substantive freedom" in that the capacity to learn effective new preferences allows the individual to achieve states of beings and doings that he himself deems valuable.[48]

Finally, our account of well-being relates to a non-standard concept of welfare that has gained much attention recently, namely, the "opportunity" criterion developed by Sugden (2004, 2007, 2008). Using a contractarian thought experiment, Sugden argues that people uncertain about the future paths of preference development will wish to have, in the future "game of life", maximum opportunities to engage in the learning of new preferences and to have these new preferences satisfied, *whatever* their content may be. For normative and formal reasons, Sugden excludes the option to engage in self-commitment (e.g. in order to overcome issues of self-control) in the domain of private good consumption. This concept of well-being overlaps with our account to the extent that both emphasize people's ability to engage in ongoing preference learning as a key constituent of well-being. Sugden's concept, however, excludes the considerations (inherent in our account) related to whether this learning is instrumentally effective (or, rather, whether it turns out to be *systematically ineffective*). Hence, it excludes considerations related to the implications, given above, of our evolutionary model of preference formation: There is no guarantee that the process of learning works perfectly well in all circumstances, as perceived by the individuals themselves. While "hard paternalistic" policies would restrict people's ability to learn new preferences (hence, decreasing well-being in our sense), our account does not rule out "libertarian paternalistic" tools which frame people's choice environments in such a way that the process of preference learning is supported.

---

[47] At a basic methodological level, there are, however, two important differences between the capability approach and our account of well-being. First, at its origins, the capability approach was motivated by the inability of the orthodox concept of welfare to cope with normative issues arising out of preference endogeneity, in particular the problem of "adaptive preference formation" (Qizilbash 2008). The solution was to define well-being in an essentially objectivist way, i.e. independent of subjective preferences (or the way they change). By contrast, our account defines well-being in more (albeit not perfectly) "subjectivist" terms. Second, our account is much more firmly grounded in empirical research on human behavior and preference formation. Notice, e.g., the conspicuous lack of interest in behavioral economics in Sen's most recent writings (e.g. Sen 2009).

[48] See Sen (1996).

This brief review of alternative non-standard notions of well-being reveals that our evolutionary account offers two additional advantages (besides the three that have been offered above). Its fourth advantage lies in the fact that it provides a relatively "complete" picture of human well-being, at least in comparison to the happiness and the opportunity account. Consider the former: The notion of happiness captures only part of what real-world individuals care about,[49] thereby running the risk of overriding people's own subjective desires and concerns. As the ongoing ability to learn new preferences is a much more fundamental element of a good life (in most people's lives) than "experiencing happiness", our account does not risk generating these "paternalistic" implications to the same extent. The same proviso applies to Sugden's notion of opportunity, due to the fact that it has to exclude people's wide-spread and perfectly natural wish to engage in self-commitment when confronted with problems of systematically dysfunctional choice and self-control. Moreover, there is evidence that only a minority of individuals share the view that welfare resides in the maximization of opportunities to satisfy any future personal preference whatsoever.[50]

Finally, a fifth advantage of our account can be argued to be found in its openness to a multi-dimensional—"pluralist"—definition of human well-being. This has been revealed by gauging the way our account relates to the three distinct "currencies" of well-being discussed above (viz., happiness, capability, and opportunity). Both the needs driving preference learning and the (temporary) outcomes of preference learning can be described as being closely related to all three dimensions of human welfare. Framing the problem of human well-being in terms of a pluralist framework adds realism to the project, considering the multi-faceted nature of human welfare that emerges from both evolutionary and behavioral research.[51]

5 Some Conceptual and Practical Implications

We conclude by qualifying the methodological status of the account of well-being proposed in Section 4 above, by presenting some preliminary thoughts on *practical* policy implications, supplemented by an exemplary *conceptual* implication relevant for welfare economics proper, and by suggesting directions for further research.

---

[49] To illustrate, all available empirical evidence shows that child-rearing has a significant negative impact on parents' happiness (see, e.g., Di Tella et al. 2003; Clark 2007; Blanchflower 2009). However, many people decide to have children and to raise them, simply on the grounds of other goals in life ("achievements", say, or the compliance with social norms).

[50] See, for example, Vis and van Kersbergen (2007) and Boeri et al. (2001). See also the related melancholic thoughts offered by Buchanan (2005).

[51] See also Sen (2009) on the implausibility of a non-pluralist notion of welfare, in particular in the context of accounts of justice.

The account of well-being that we suggest is but a first step towards a fully developed evolutionary account of welfare. It provides a general framework to guide and to structure the development of specific normative arguments and policy scenarios. More specifically, it may refocus the citizens' and policy-makers' attention to certain (possibly hitherto neglected) problems that are particularly relevant for a smooth working of evolutionary processes of change, when considered from a micro perspective. There is obviously still a long way to go until we have a conceptual framework that allows us to derive policy implications that are both fully consistent with the background assumptions of an evolutionary world-view and perfectly operational. Even then, however, it is worth recalling the prediction made by Nelson and Winter (1982: 357) in their plea for a rethinking of welfare economics: "[A] normative theory consistent with an evolutionary approach to positive theory almost certainly will be *complex and messy*. It is unlikely that one will be able to prove many sweeping normative theorems of the sort that are now contained in our advanced treatises and elementary texts. This, however, should not cause despair."[52] From an evolutionary perspective, it is hard to imagine otherwise, since "one important consequence of the Darwinian theory was to banish the idea of perfection from the discussion of progress" (Metcalfe 2001: 565). A theory that would allow us to derive a complete ordering of a given vector of social states (and, a fortiori, processes) is clearly out of sight.[53] In an evolutionary setting, rather than identifying "optimal" institutions, we are left with the (pragmatically more important) task of the comparative analysis of alternative arrangements. It is reasonable to assume that, ultimately, welfare consists of several components which are not necessarily reducible to some underlying one-dimensional "basic currency" and basic measuring rod. Accordingly, the assessment of an agent's motivation and ability to engage in effective preference learning may be done from a variety of different angles. While further research may unveil techniques to operationalize and to weigh these angles, we suggest that, already in its present preliminary state, our account of well-being may provide citizens and policy-makers with some valuable heuristics about how to think about normative issues (in particular the problem of compensation) brought about by processes of "creative destruction".

According to our account of well-being, innovation is worth promoting—and the associated risks are worth enduring—to the extent that it contributes to an improvement in people's ability to explore new preferences over time. The policy focus shifts from the level of given preferences and their "rational" satisfaction to the deeper level of the factors determining and the conditions facilitating the formation of preferences. To the extent that policy (i.e., rule design) is able to influence those institutional conditions, it should be guided by two basic considerations:

– First, it should maintain and enlarge people's freedom to choose among different paths of preference learning. To illustrate, the scope of practices restricting the

[52] My italics.

[53] On the inevitable incompleteness of evaluations in matters of social justice, see Sen (2009: 103–08).

variety of markets (such as monopolies, tariffs, quotas, or discriminatory labor laws) should be reduced.
– Second, policy should maintain the institutional preconditions of individuals' ability to learn and to explore new preferences by providing appropriate "choice environments".

In light of these two basic requirements, three sets of policy recommendations come to mind. First, policy should focus on the quality of the institutional framework underlying people's "choice environment" (Anand and Gray 2009). In this context, a set of natural candidates for policy instruments is provided by the toolbox of libertarian paternalism. By framing the informational context of economic choices and by designing default rules, so-called "nudges" have been shown to have a significant impact on individual behavior, e.g. with respect to issues such as overborrowing, insufficient retirement saving, risk-taking, status consumption (and ensuing "status races") or health. What these tools have in common is that they steer people's behavior in a way that is in their own best interests (as indicated by levels of ex post regret), while at the same time leaving opportunity sets essentially unchanged. It is conceivable that these tools can be applied in settings in which people's welfare, as understood in our account, is low. To start with the most obvious risk to people's ability to engage in ongoing preference learning, the consumption of *addictive substances* may be targeted by issuing and prescribing appropriate warnings. As a second example, consider the issue of excessively "borrowing-friendly" environments (Anand and Gray 2009), where the ease of obtaining debt induces poorly informed agents into behavioral cycles that they have reason to avoid ex ante. This tendency may be exacerbated by systematic behavioral "biases" in favor of, e.g., mispredicting one's future utility by overestimating the weight of immediate gratification relative to deferred well-being. A final example concerns the consumption of status-signaling goods (Frank 2008) which may lead people into interdependent status races that are "wasteful" to the extent that they may be absorbed in self-augmenting cycles of preference acquisition and consumption that may, ultimately, undermine the process of learning itself (Cordes and Schubert 2010). Libertarian paternalist tools may help limit people's inclination to engage in these kinds of consumption cycles.[54] To generalize, issues related to addiction, overborrowing, and status races may be normatively relevant (in terms of our account of well-being) to the extent that it can be shown that they reflect systematic "biases" of individuals to engage in self-

[54] Note that the evolutionary model of preference formation discussed in Section 4.1 above, may back the following qualification of "libertarian paternalist" tools. According to the model, agents acquire preferences on a non-cognitive and, based on this, on a "higher" cognitive level. Given that the former provides the essential building blocks ("needs") for the subsequent refinement achieved on the cognitive level, one may argue that libertarian paternalism should refrain from using any tool that affects non-cognitive learning processes, i.e., that cannot be made subject to conscious deliberation at the level of cognitive reasoning. In this way, our evolutionary model would help in providing an explicit justification to an argument that is usually made ad hoc; see, e.g., Bovens (2008).

defeating behavioral patterns. In that case, it may be legitimate to channel the learning dynamics away from "destructive" behavioral equilibria.

A second set of policy recommendations concerns innovation policy, perceived broadly as including any kind of policy that has a positive impact on the intensity and direction of innovative (in particular entrepreneurial) activity. At a most basic level, our account of well-being encourages policy to invest in education and training at all levels of the economy. It seems that well-being in the sense defined here could be fostered particularly effectively by investing in the education of children about how to engage in "second-order learning", i.e. the ability to learn (Hodgson 1999). Fostering entrepreneurial activity by, e.g., improving the access of innovative small firms to venture capital, fits easily into our normative framework as well.

Finally, a third set of practical implications refers to the welfare state that is characteristic of advanced economies. There is a widespread intuition that, in the way it is typically organized, the welfare state may have a counter-productive impact on the well-being, properly conceived, of welfare recipients. This is most visible from a dynamic perspective. Unconditional welfare transfers may, for instance, discourage entrepreneurial activity and risk-taking (Henrekson 2005). Behavioral economics is able to make sense of this by rationalizing, e.g., the effect of transfers on people's partly "innate" proclivity to underestimate the present value of future investment-induced benefits, or on their willingness to assume responsibility for their own life (Beaulier and Caplan 2007). Accordingly, an account of well-being that is informed by empirical evidence on human behavior can provide good reasons for refocusing the welfare state's toolbox. In an evolving economy, the welfare state should avoid discouraging people's motivation and ability actively to participate in the learning economy, and instead provide for a basic degree of stability and security to encourage risk-taking and explorative action.

As has been outlined in 'Section 1' above, at a more abstract, conceptual level, an evolutionary approach to welfare may also enrich welfare economics proper. More specifically, it may offer a fresh perspective on some thorny issues related to preference change. To illustrate this claim, consider the problem of adaptive preference change (Elster 1982) that is relevant, e.g., in policy issues pertaining to developing economies. People situated in "objectively" miserable circumstances may learn to accept their situation to the point that they no longer prefer anything that is not available to them—the "hopeless beggar, the precarious landless laborer" and the "dominated housewife" are cases in point (Sen 1988: 45–46). From an orthodox viewpoint, their manifest preferences and the degree of their "satisfaction" would indicate rather high levels of well-being. This, however, is counterintuitive. While Sen concludes that purely subjectivist accounts of welfare should be abandoned in favor of the more objective account of capability, others have tried to modify accounts of preference-based welfare in such a way as to accommodate for the "adaptation problem" (Qizilbash 2008). This has turned out to be difficult, though: Rescuing operations involving so-called full-information accounts of welfare have failed, for the individuals displaying adaptive preference formation may very well be perfectly aware of their situation (Sobel 1994). The account of welfare suggested in the present paper may be helpful here: Sen's individuals may show a comparatively low motivation and ability to engage in instrumentally effective preference learning.

As demonstrated over the past three decades, evolutionary economics has far-reaching implications for the way economists theorize about individual behavior and social phenomena in an evolving market economy. The present paper has demonstrated that it may also require a fundamental re-orientation of welfare economics, in particular, with respect to the normative issues related to endogenous preference change. More specifically, innovation-driven change may have normatively relevant side-effects to the extent that it deprives individuals from the chance actively to participate in the learning economy. Obviously, much further research is required to back this intuition with analytical means. For instance, we need to know more about the aversive effects that increasing innovativeness and competitiveness have on the well-being of real-world individuals and on their inclination and willingness to be explorative and creative. Another desideratum are clear criteria for judging whether processes of preference formation are prone to self-defeating ("disequilibrating", Cordes and Schubert 2010) dynamics. Finally, we need criteria to clarify to what extent insights from the neighboring fields of behavioral economics and psychology proper may be used to refine our account of well-being.

Ultimately, an operational account of welfare will enable evolutionary economists to back their (conditional) support for the innovation-driven, experimental economy by explicit reference to consistent welfare criteria, implying that the purpose of economic activity is not innovation, but individual well-being. Innovation is worth pursuing to the extent that it promotes well-being, properly understood. There are good reasons to believe that referring to such a standard (attesting to the economists' sensitivity to the ethical dimension of economic change) will increase the chances that the evolving economy is perceived as legitimate.

References

Aldrich HE, Hodgson GM, Hull DL, Knudsen T, Mokyr J, Vanberg VJ (2008) In defence of generalized Darwinism. J Evol Econ 18:577–596

Anand P, Gray A (2009) Obesity as market failure: could a 'deliberative economy' overcome the problem of paternalism? Kyklos 62:182–190

Andersen ES (2004) Population thinking and evolutionary economic analysis: exploring Marshall's fable of the trees. Mimeo

Ariely D, Loewenstein G, Prelec D (2006) Tom Sawyer and the construction of value. J Econ Behav Organ 60:1–10

Atkinson AB (2009) Economics as a moral science. Economica 76:791–804

Audretsch DB, Grillo I, Thurik AR (2007) Explaining entrepreneurship and the role of policy: a framework. In: Handbook of research in entrepreneurship policy. Elgar, Cheltenham, pp 1–17

Baumol W (1990) Entrepreneurship: productive, unproductive, and destructive. J Polit Econ 98:893–921

Beaulier S, Caplan B (2007) Behavioral economics and perverse effects of the welfare state. Kyklos 60:485–507

Bianchi M (2002) Novelty, preferences, and fashion: when goods are unsettling. J Econ Behav Organ 47:1–18
Binder M (2010) Elements of an evolutionary theory of welfare. Routledge, London
Blanchflower DG (2009) International evidence on well-being. In: Kruger AB (ed) Measuring the subjective well-being of nations: national accounts of time use and well-being. Chicago University Press, Chicago, pp 155–226
Blanchflower DG, Oswald AJ (2005) Some policy implications of behavioral economics–Happiness and the human development index: the paradox of Australia. Aust Econ Rev 38:307–318
Boeri T, Börsch-Supan A, Tabellini G (2001) Would you like to shrink the welfare state? A survey of European citizens. Econ Pol 16:9–50
Boulding KE (1981) Evolutionary economics. Sage, Beverly Hills
Bovens L (2008) The ethics of Nudge. In: Grüne-Yanoff T, Hansson SO (eds) Preference change: approaches from philosophy, economics and psychology. Springer, Berlin, pp 207–220
Broome J (1996) Choice and value in economics. In: Hamlin AP (ed) Ethics and economics, vol. I. Elgar, Cheltenham, pp 65–85
Broome J (2008) Why economics needs ethical theory. In: Basu K, Kanbur R (eds) Arguments for a better world: essays in honor of Amartya Sen. Oxford University Press, New York, pp 7–15
Bruni L, Sugden R (2007) The road not taken: how psychology was removed from economics, and how it might be brought back. Econ J 117:146–173
Buchanan JM (1977) Freedom in constitutional contract. Texas A & M University Press, College Station
Buchanan JM (2005) Afraid to be free: dependency as desideratum. Public Choice 124:19–31
Camerer C, Issacharoff S, Loewenstein G, O'Donoghue T, Rabin M (2003) Regulation for conservatives: behavioral economics and the case for 'asymmetric paternalism'. Univ PA Law Rev 151:1211–1254
Cantner U, Pyka A (2001) Classifying technology policy from an evolutionary perspective. Res Policy 30:759–775
Clark DA (2006) The capability approach: its development, critiques and recent advances. In: Clark DA (ed) The Elgar Companion to development studies. Edward Elgar, Cheltenham, pp 32–45
Clark AE (2007) Born to be mild? Cohort effects don't (fully) explain why well-being is U-shaped in age. IZA Discussion Paper No. 3170, IZA, Bonn, Germany
Cordes C (2008) A potential limit on competition. J Bioecon 10:127–144
Cordes C, Schubert C (2010) Role models that make you unhappy: light paternalism, social learning and welfare. Papers on Economics & Evolution #1022. Max Planck Institute of Economics, Jena
Di Tella R, McCulloch R, Oswald A (2003) The macroeconomics of happiness. Rev Econ Stat 85:809–827
Dopfer K (1976) Introduction: towards a new paradigm. In: Dopfer K (ed) Economics in the future. Macmillan, London, pp 3–35
Dosi G, Marengo L, Pasquali C (2006) How much should society fuel the greed of innovators? On the relations between appropriability, opportunities and rates of innovation. Res Policy 35:1110–1121
Earl PE, Potts J (2004) The market for preferences. Camb J Econ 28:619–633
Elster J (1982) Sour grapes–utilitarianism and the genesis of wants. In: Sen AK, Williams B (eds) Utilitarianism and beyond. Cambridge University Press, Cambridge, pp 219–238
Foss NJ (2006) Evolutionary Economics and Economic Policy. Available online at: http://organizationsandmarkets.com/2006/08/02/evolutionary-economics-and-economic-policy/
Frank RH (2008) Should public policy respond to positional externalities? J Public Econ 92:1777–1786
Frey BS, Stutzer A (2002) What can economists learn from happiness research? J Econ Lit 60:402–435

Frey BS, Sutzer A (2007) What happiness research can tell us about self-control problems and utility mispredictions. In: Frey BS, Stutzer A (eds) Economics and psychology: a promising new cross-disciplinary field. MIT Press, Cambridge, pp 169–195
Frey BS, Benz M, Stutzer A (2004) Introducing procedural utility: not only what, but also how matters. J Inst Theor Econ 160:377–401
Georgescu-Roegen N (1954) Choice, expectations, and measurability. Q J Econ 68:503–534
Gowdy JM (2004) The revolution in welfare economics and its implications for environmental valuation and policy. Land Econ 80:239–257
Gray J (1999) Hayek on liberty. Routledge, London
Hanusch A, Pyka A (2007) Principles of Neo-Schumpeterian economics. Camb J Econ 31:275–289
Harsanyi J (1982) Morality and the theory of rational behavior. In: Sen AK, Williams B (eds) Utilitarianism and beyond. Cambridge University Press, Cambridge, pp 39–62
Hayek FA (1948) Individualism and economic order. University of Chicago Press, Chicago
Hayek FA (1976) Law, legislation and liberty, vol. II. University of Chicago Press, Chicago
Hayek FA (1978) The confusion of language in political thought. In: New studies in philosophy, politics, economics and the history of ideas. Routledge, London, pp 71–97
Hayek FA (1988) The fatal conceit. University of Chicago Press, Chicago
Hayek FA (2009) The common sense of progress. In: The constitution of liberty. Routledge, London, pp 36–48
Henrekson M (2005) Entrepreneurship: a weak link in the welfare state? Ind Corp Change 14:437–467
Herrnstein RJ (1990) Behavior, reinforcement and utility. Psychological Science 1:217–224
Hodgson GM (1993) Economics and evolution: bringing life back into economics. University of Michigan Press, Ann Arbor
Hodgson GM (1999) Economics and Utopia: why the learning economy is not the end of history. Routledge, London
Ironmonger DS (1972) New commodities and consumer behavior. Cambridge University Press, Cambridge
Kahneman D, Wakker PP, Sarin R (1997) Back to Bentham? Explorations of experienced utility. Q J Econ 112:375–405
Kahneman D, Sugden R (2005) Experienced utility as a standard of policy evaluation. Environ Resource Econ 32:161–181
Kerstenetzky CL (2007) Hayek and Popper on ignorance and intervention. J Inst Econ 3:33–53
Keynes JN (1917) The scope and method of political economy. Macmillan, London
Köszegi B, Rabin M (2008) Choices, situations, and happiness. J Public Econ 92:1821–1832
Lachmann L (2007) Capital and its structure. Sheed Andres and McMeel, Kansas City
Lessig L (1995) The regulation of social meaning. Univ Chic Law Rev 62:943–1045
Little IMD (1957) A critique of welfare economics. Oxford University Press, London
Metcalfe JS (2001) Institutions and progress. Ind Corp Change 10: 561–586
Metcalfe JS (2005) Systems failure and the case for innovation policy. In: Llerena P, Matt M (eds) Innovation policy in a knowledge-based economy. Springer, Berlin, pp 47–74
Mokyr J (2000) Innovation and its enemies: the economic and political roots of technological inertia. In: Olson M, Kähkönen S (eds) A not-so-dismal science. Oxford University Press, Oxford, pp 61–91
Myrdal G (1933) Das Zweck-Mittel Denken in der Nationalökonomie. Zeitschrift für Nationalökonomie 4:305–329
Nelson RR (1977) The Moon and the Ghetto. Norton, New York
Nelson RR (1981) Assessing private enterprise: an exegesis of tangled doctrine. Bell J Econ 12:93–111
Nelson RR (1990) Capitalism as an engine of growth. Res Policy 19:193–214
Nelson RR (1995) Recent evolutionary theorizing about economic change. J Econ Lit 33:48–90

Nelson RR, Winter SG (1982) An Evolutionary Theory of Economic Change. Belknap Press, Cambridge
Ng Y-K (2003) From preference to happiness: towards a more complete welfare economics. Soc Choice Welf 20:307–350
North DC (1990) Institutions, institutional change and economic performance. Cambridge University Press, Cambridge
Nozick R (1974) Anarchy, state and utopia. Basic Books, New York
Ostrom E (2005) Understanding institutional diversity. Princeton University Press
Pelikan P (2002) Why economic policies need comprehensive evolutionary analysis. In: Pelikan P, Wegner G (eds) Evolutionary thinking on economic policy. Elgar, Cheltenham, pp 15–45
Potts J (2001) Knowledge and markets. J Evol Econ 11:413–431
Potts J (2004) Liberty bubbles. Policy 20:15–21
Qizilbash M (2008) The adaptation problem, evolution and normative economics. Papers on Economics & Evolution #0708. Max Planck Institute of Economics, Jena
Rabin M (2002) A perspective on psychology and economics. Eur Econ Rev 46:657–685
Rawls J (1971) A Theory of Justice. Belknap Press, Cambridge
Robbins L (1935) An Essay on the Nature and Significance of Economic Science. Macmillan, London
Samuelson PA (1938) A note on the pure theory of consumer's behaviour. Economica 5:61–71
Sartorius C (2003) An Evolutionary Approach to Social Welfare. Routledge, London
Scanlon TM (1991) The moral basis of interpersonal comparisons. In: Elster J, Roemer JE (eds) Interpersonal comparisons of well-being. Cambridge University Press, MA, pp 17–44
Schubert C (2009) Welfare Creation and Destruction in a Schumpeterian World. Papers on Economics & Evolution #0913. Max Planck Institute of Economics, Jena
Schumpeter JA (1912) Theorie der wirtschaftlichen Entwicklung. Duncker & Humblot, Berlin
Schumpeter JA (1942) Capitalism, Socialism and Democracy. Allen & Unwin, London
Schumpeter JA (1954) History of Economic Analysis. Allen & Unwin, London
Scitovsky T (1941) A note on welfare propositions in economics. Rev Econ Stud 9:77–88
Sen AK (1977a) Rational fools: a critique of the behavioral foundations of economic theory. Philos Public Aff 6:317–344
Sen AK (1977b) On weights and measures: informational constraints in social welfare analysis. Econometrica 45:1539–1572
Sen AK (1979) Equality of What? Tanner Lecture on Human Values. Available online at: http://www.tannerlectures.utah.edu/lectures/documents/sen80.pdf
Sen AK (1980) Equality of what. In: McMurrin SM (ed) The tanner lectures on human values. University of Utah Press, Salt Lake City, pp 197–220
Sen AK (1985) Commodities and Capabilities. Basic Books, New York
Sen AK (1988) On Ethics and Economics. Basil Blackwell, Oxford
Sen AK (1993) On the Darwinian view of progress. Popul Dev Rev 19:123–137
Sen AK (1996) On the foundations of welfare economics: utility, capability, and practical reason. In: Farina F, Hahn F, Vanucci S (eds) Ethics, rationality, and economic behavior. Clarendon, Oxford, pp 50–65
Sen AK (1999) Development as Freedom. Knopf, New York
Sen AK (2009) The Idea of Justice. Allen Lane, London
Sobel D (1994) Full-information accounts of well-being. Ethics 104:784–810
Stam E (2008) Entrepreneurship and innovation policy. Jena Economic Research Papers #2008-006
Sugden R (1993) Normative judgments and spontaneous order: the contractarian element in Hayek's thought. Const Polit Econ 4:393–424
Sugden R (2004) The opportunity criterion: consumer sovereignty without the assumption of coherent preferences. Am Econ Rev 94:1014–1033
Sugden R (2007) The value of opportunities over time when preferences are unstable. Soc Choice Welf 29:665–682

Sugden R (2008) Why incoherent preferences do not justify paternalism. Const Polit Econ 19:226–248
Sugden R (2009) On nudging: a review of nudge: improving decisions about health, wealth and happiness. Int J Econ Bus 16:365–373
Sunstein CR, Thaler RH (2003) Libertarian paternalism. Am Econ Rev, Papers & Proceedings 93:175–179
Thaler RH, Sunstein CR (2008) Nudge: improving decisions about health, wealth, and happiness. Princeton University Press, New Haven
Van den Bergh JCJM, Kallis G (2009) Evolutionary policy. Papers on economics & evolution #0902. Max Planck Institute of Economics, Jena
Van Praag M, Versloot PH (2007) What is the value of entrepreneurship? A review of recent research. Small Bus Econ 29:351–382
Vanberg VJ (1994a) Cultural evolution, collective learning, and constitutional design. In: Reisman D (ed) Economic thought and political theory. Kluwer, Boston, pp 171–204
Vanberg VJ (1994b) Hayek's legacy and the future of liberal thought: rational liberalism vs. evolutionary agnosticism. J des Econ et des Etud Hum 5:451–481
Vanberg VJ (2006) Human intentionality and design in cultural evolution. In: Schubert C, von Wangenheim G (eds) Evolution and design of institutions. Routledge, London, pp 197–212
Vis B, van Kersbergen K (2007) Why and how do political actors pursue risky reforms? J Theor Polit 19:153–172
Wegner G (1997) Economic policy from an evolutionary perspective–a new approach. J Inst Theor Econ 153:485–509
Wegner G (2005) Reconciling evolutionary economics with liberalism. In: Dopfer K (ed) Economics, evolution and the state. Elgar, Cheltenham, pp 58–77
von Weizsäcker CC (1971) Notes on endogenous change of tastes. J Econ Theory 3:345–372
von Weizsäcker CC (2010) Cost-Benefit Analysis with Adaptive Preferences. Mimeo, Max Planck Institute for Research on Collective Goods, Bonn, Germany
Whitman DG (1998) Hayek contra pangloss on evolutionary systems. Const Polit Econ 9:45–66
Witt U (1987) Individualistische Grundlagen der evolutorischen Ökonomik. Mohr Siebeck, Tübingen
Witt U (1996) Innovations, externalities and the problem of economic progress. Public Choice 89:113–130
Witt U (2000) Genes, culture, and utility. Papers on Economics & Evolution #0009. Max Planck Institute of Economics, Jena
Witt U (2001) Learning to consume–a theory of wants and the growth of demand. J Evol Econ 11:23–36
Witt U (2003) Economic policy making in evolutionary perspective. J Evol Econ 13:77–94
Witt U (2004) Beharrung und Wandel – ist wirtschaftliche Evolution theoriefähig? Erwägen – Wissen – Ethik 15:33–45
Witt U (2008) What is specific about evolutionary economics? J Evol Econ 18:547–575
Witt U, Schubert C (2010) Extending the informational basis of welfare economics–the case of preference dynamics. Papers on Economics & Evolution #1005. Max Planck Institute of Economics, Jena

Part III
Innovation and the Evolution of Capitalism

Dynamic Circular Flow Models with Innovations

Some Methodological Aspects of the Banking and Finance Crisis from the Point of View of Evolutionary Economics

Adolf Wagner

Abstract The paper starts with principle assertions in evolutionary economics. On the one hand, the consistent microfoundation of economic aggregates is mathematical impossible, and on the other hand, any representative microfoundation is not acceptable for evolutionary macroeconomics. So we have to work with the assumption of some "aggregative stability" over a not too long period of time. In a further step the concepts of traditional flow equilibria and the concepts of circulation disequilibrium are outlined. Here, circulation disequilibria and changing stocks of money are the normality following the research tradition of Tuebingen University. Of special interest are the so called "Saldenmechanik" as well as the "Maximalbelastungsrechnung". Only institutional transactors are relevant, for only these sectors manage stocks and flows. Possible innovations are evident: there might be new sectors and new transactions. We end with some conclusions and a historical note.

1 On Some Characteristics of Evolutionary Macroeconomics

1.1 Economics with Broad Economic Aggregates

"The term 'macro-economics' introduced by Ragnar Frisch in 1933, applies to the study of relations between broad economic aggregates, as opposed to the decision-making processes of individuals and firms which is the subject-matter of micro-economics" (Allen 1968, p. 1). Some theories of political economics are macroeconomic from the beginning, e.g. business cycle theory, foreign trade theory and especially circular flow models (Schlicht 1990, p. 287). This has to do with the

A. Wagner (✉)
Leipzig University, Burglehenweg 7, 72108 Rottenburg, Germany
e-mail: prof@adolfwagner.eu

G. Buenstorf et al. (eds.), *The Two Sides of Innovation*, Economic Complexity and Evolution, DOI 10.1007/978-3-319-01496-8_12,

mathematical impossibility of consistent microfoundations as well as with the following aspect.

1.2 We Always Have to Deal with Principally Inhomogenous Aggregates

Aggregates or the aggregate variables represent the sums or the averages of principally very heterogenous items of individuals and firms (see Krelle 1978, p. 1). There is no such thing as a "perfect aggregation" from the micro to the macro level (see Allen 1965, p. 719, and Pokropp 1972). We always have to rely on the assumption of some "aggregative stability" (see definition in Wagner 1983, p. 441) within some periods of analysis and for some questions.

1.3 An Illustrative Picture of "Population Thinking" and Structural Change

In honor of Alfred E. Ott, Tycho Seitz presented a paper in 1989 with an illustrative picture of "population thinking" (already founded in the elder heterodox population economics), often mentioned in evolutionary economics: "The fundamental reasons for the permanent structural change in the economy certainly lie in the demographic processes of global ageing. Even with stationary populations during the elapse of time we see ever changing collections of human beings making decisions, gathering experiences and findings. At the same time men lose experiences and findings from former times. History leads to a difficult balance of learning and forgetting. So the ideas of life and the attitudes against fellow-men and problems are ever changing. Quite exceptionally all aspects of individual and collective life are permanently changing" (Seitz 1989, p. 35). These aspects become even more important within growing or shrinking populations.

1.4 Three Well-Known Misjudgments in Traditional Macroeconomics

First of all, we have to abstain from all kinds of "representative microfoundations" of aggregates in evolutionary macroeconomics. This concerns the "average man" (or "homme moyen") of the statistician Adolphe Quetelet (1796–1874) as well as the opinion of Rudolf Richter, that macroeconomics could be a handy form of microeconomics (Richter 1979, p. 173). Second, we have to ignore the view that equilibrium in macroeconomics always is realized on macro markets and macro market equilibrium (see Ramser 1987, p. 38). Instead, we have to use the concept of "circulation equilibrium" in the case of fulfilled budget equations for all macroeconomic transactors or sectors (Wagner 1989, 2003, pp. 510–516). Third, one has to abstain from the occasionally fatal mixtures of flow equilibria and planning equilibria. For example: "Any IS/LM equilibrium is a momentary equilibrium that is liable to be disrupted both for reasons relating to the disappointment of expectations and for reasons stemming from stock-flow inconsistencies" (Metcalfe and Steedman 1991, p. 77).

2 Traditional Flow Equilibria

2.1 Representations of Circular Flow in Macroeconomics

There are four kinds of representation of circular flow in macroeconomics (Wagner 2009, pp. 61–66): (1) circular-flow diagrams, (2) matrix-representations, (3) representations by budget-equations, (4) T-account-representations. Each of them consists of two elements: (a) transactors or sectors (e. g. caskets or circles, rows, titles in columns and rows, headlines of accounts); (b) transactions or streams (i. e. arrows, items, entries). International textbooks (see Mankiw and Taylor 2006, p. 23 and 488, Krugman and Wells 2006, pp. 31–32) first and foremost make use of (1). In some cases, we have to distinguish institutional sectors (e. g. consumers, investors, enterprises, governments) and functional sectors (e.g. markets, income accounts). In an extremely simple and special case, one can say: "The circular-flow diagram is a visual model of the economy that shows how money and production inputs and outputs [streams of money and goods] flow through markets [functional sectors] among households and firms [sectors]" (Mankiw and Taylor 2006, p. 23).

2.2 The Concept of Circulation Equilibrium or Flow Equilibrium

"Circulation equilibrium" is realized in a period of analysis, if and only if every sector or transactor shows equal sums of receipts (inputs) and expenditures (outputs). The result of circulation equilibrium is that there are constant monetary stocks of fortunes at the end and at the beginning of a period in every transactor or sector. Therefore, the net of debts and claims or credits remains unchanged in the case of circulation equilibrium. With equal steps in transactions and circulation equilibrium, traditional banking would be superfluous. We can conclude: Circulation equilibrium leads to sustainability (with constant stocks at all sectors). A special and often misunderstood topic is "Walras law" (after Léon Walras, 1834–1910): If one is told that a circulation system of n sectors has fulfilled budget equations with n-1 sectors, the remaining sector must also be in input–output balance (but Walras law is not the Walras system). In contrast, each circulation disequilibrium (and unequal or uncoordinated steps of transactions) requires a certain management of positive or negative sectoral balances by banks. Here we see logical connections to the present financial crisis.

2.3 Book-Keeping Systems with Built-In Circulation Equilibria

A profound misunderstanding prevails between a true circular flow model, on the one hand (with the possibility of equilibrium and disequilbrium), and a mere book-keeping system with built-in circulation equilibrium (for all economic conditions), on the other hand. Prominent examples of book-keeping systems are all the systems of national accounts (p. e. United Nations 1993). The national accounts are sometimes taken falsely for true circular flow equilibria of an economy (see Krugman and Wells 2006, p. 839). Therefore, it is important to point out the specifica of book-keeping systems. For every budget-equation, there comes along a specific balancing "variable" and at last the above mentioned "Walras law". Two kinds of specific balancing variables are used for the sectoral budget-equations: sum variables and balancing variables in a narrow sense. Examples: changes of monetary wealth in the household sector (identical with income minus expenditures), or savings identical with income minus consumption in the household sector, income identical with consumption plus investment in the enterprise sector (see Wagner 2009, pp. 64–65). Within the system of national accounts, circulation equilibria always hold in a formal sense. Every statistical yearbook showsthat investment equals savings. But in the real world and its circular flow models, it remains a question whether planned investment of the enterprise sector in total equals voluntary savings of the household sector.

3 Dynamic Concepts of Circulation Disequilibrium

3.1 Circulation Disequilibria and Changing Stocks of Money Are the Normality

Hans Peter (1898–1959) was convinced that the budget equations of the transactors in general are not empirically valid and that, therefore, as a rule, circulation leads to responses in the credit net. Huge unmanageable differences might in principle represent a financial crisis. Wolfgang Stützel (1925–1987)—a professor with solid experience in the banking business—wondered about the frequent use of balanced budgets in economic theory (see Helmstädter 1981, p. 155, Stützel 1978, p. 262). Yet, normality in macroeconomics means circulation disequilibria, period by period changing stocks of positive or negative wealth at the transactors, and stress on the banking system with deposits and loans.

3.2 Disequilibrium Flow Economics and the "Saldenmechanik"

By his "Saldenmechanik" (see Stützel 1978), Wolfgang Stützel aimed at the system of macroeconomic budget equations and the corresponding difference equations according to the national accounting system—regardless of certain empirically valid behavioral, institutional or technological equations, and regardless of equilibrium or disequilibrium. Stützel intended to prevent the economists from premature and unrestricted theorizing without seeing the boundaries of the accounting identities. An important case of applying the "Saldenmechanik" persists in the stocks of money or wealth, especially the money of transfer ("Buchgeld") and its reaction on circulation disequilibrium.

3.3 Dynamic Models of Flow Disequilibrium Are Needed

Dynamic models of flow disequilibrium need further research, as we can learn from the ongoing banking and finance crisis. The aspects of circulation disequilibrium were neglected in the econometric macroeconomic modelling along with the Cowles-Commission-Type models in the past. Perhaps one finds some starting-points within simple non-mathematical approaches of Karl-Heinz Raabe (see Raabe 1967, 1974), who used tables of national accounting to fill in the items for present and future periods. He hereby followed the principles of the "Saldenmechanik". As I remember, Raabe began with "mass incomes" and then went to consumption in the following period and other aggregates of minor importance. With his calculations he aimed at a best fit with the accounting results of the period coming. The empirical-statistical proceeding of Raabe follows certain ideas of the elder growth and business cycle theories (see Ihring 1998; Reichardt 1967; Peter and Reichardt 1959).

3.4 "Maximalbelastungsrechnung" and System Crises in Flow Economics

We already saw with my famous namesake Adolph Wagner (1835–1917) the necessity for every bank to always remain sound and solvent but nevertheless earn money by lending out money. Thus he formulated a so-called "golden banking rule" in 1857 (see Wagner 1857, p. 167 on). The "golden rule of finance" or the "golden banking rule" means the following: The granted loans and credits on the active side of the balance sheet should correspond to the deposits on the passive side of the balance sheet with regard to volumes, spaces of time and payment schedules. On the occasion of a "run" on a British bank, Wolfgang Stützel (1925–1987) in 1959 reviewed the so-called "golden banking rule". His central question was: What must be the trading and risk policy for a bank so that the deposits remain sure at any moment (see Stützel 1959, pp. 35–36)? The "golden banking rule" as well as the "sediment theory" ("Bodensatztheorie") only hold in "normal times" without any panic and with full operating markets. During times of crisis—that was the leading idea—everything is different to realize. All valuations decay dramatically. By his "Maximalbelastungsrechnung"—a maximum stress analysis—Wolfgang Stützel found a real "stress test" about 50 years before the word "stress test" became fashionable. Nowadays, journalists and accountants sometimes speak of a "liquidity-catastrophe plan" without any knowledge of Stützel. To prevent a systems crisis has been the aim of the Stützel-"Maximalbelastungsrechnung" (see Schmidt, Ketzel, Prigge 2001, p. VIII). The "Maximalbelastungsrechnung" is—so to say—the "golden banking rule" for times of crisis. From a liquidity disaster or catastrophe à la Stützel, we come to systems innovations.

4 Innovations Within Dynamic Circulation Models

4.1 Transactions-Innovations: Spread Effects and New Financial Products

Money flows from transactor to transactor depend on the substance of money. In Europe, we have bank notes and coins circulating, but a very elusive huge volume of money of transfer ("Buchgeld") as well. Whereas bank notes and coins can be measured and controlled by the central bank (except some leakages into foreign countries), the components of transfer-money vary (a) by definitions (see the occasionaly changing conventional definitions for M1, M2 and M3) as well as (b) by private arrangements within structuring money wealth. The latter activities cause so-called spread effects ("Spreizeffekte", see Stützel 1978, p. 210, 224, 227, and Wagner 2009, p. 135), so that the total amount of money in the economy partly depends on private arrangements, and not on monetary policy decisions alone.

The spread effects in transfer-money even lead to a special version of the quantity-formula of money. I restrict myself to a single case of the spread effects for illustration purposes: Money M1 partly consists of deposits. If a transaction goes from a deposit-account to a credit-account, Money M1 is shrinking (and vice versa). An earlier example for transactions-innovations are "Repo-contracts" (contracts concerning selling as well as repurchasing of securities, "Wertpapierpensions-geschäfte"), which once came in as a component of M3. In principle, some new financial products may alter the net of transactions as soon as they find application for transactions. Money is never perfectly under the control of the central banks. There is no sharp border line between money and non-money (see Kaldor and Trevithick 1981, p. 418).

4.2 Transactors-Innovations: Founding and Annulment-Effects

Innovations in circulation flow systems do not occur merely by altering the medium of circulation or exchange, but also by the addition of new transactors as well as by the abolition of old ones. If we had a descriptive picture of the net of money-streams in the economy, we could make use of input–output mathematics. The supply oriented Leontief-model is of special interest here (see Wagner 2009, p. 95). We could look for strategic sectors of circulation (see Statistisches Amt der Europäischen Gemeinschaften 1976, p. 29), we could make out the multipliers matrix and thereby identify the so-called powers of dispersion and the sensitivities of dispersion of the certain transactors. Here we look at the establishment of new enterprises and at the ruin of old ones. Newcomers from abroad also play a role, especially during the so-called globalization process. In cases of ruin, we calculate the annulment effect.

4.3 Are There Two Versions of Financial Crises?

In my opinion, crises are caused either (a) from the transactions side or (b) from the transactors' side. Case (a) has heavy disturbed budget-equations as background, leading to extraordinary biases of the stocks of wealth (positive and negative). Case (b) starts from extremely high national debts (see Greece or Ireland today), and forces for the time unmanageable transactions. Sometimes derivatives and other positions designated for "bad banks" play a role, too.

4.4 The Main Point of the Financial Crisis: Money of Transfer at Risk

To prevent bank and financial crises in the future, we need a much better behavioral codex from the bankers' side. As in former European times, banks must have a look at the macroeconomic consequences of their individual rent-seeking behavior. Each banker should be forced to a daily individual "stress test" in the sense of the Stützel-Maximalbelastungsrechnung from 1959 (see Sect. 3.4). The economic process of money-creation ("Buchgeldschöpfung") must come to the safe side again. It is not only a matter of the legal system to secure our money, which doesn't consist of banknotes and coins alone.

5 Conclusions

Summarizing the discussion leads to the following five main results:

(1) Traditional circulation or flow equilibria are perhaps the only macroeconomic equilibrium conceptions with relevance for evolutionary economics.
(2) Money flow and circulation disequilibrium should be much better investigated in the face of present and coming banking and finance crises.
(3) All real "stress tests" historically begin with Wolfgang Stützel and his 1959 "Maximalbelastungsrechnung" for daily calculations. Executives in the banks should internalize this again.
(4) Thus, we make sure that our non-cash money is no longer endangered by banking and finance crises.
(5) A marginal note: By joining the Euro, national money and exchange-rate policies were given away (problematic from the point of view of the Tinbergen-rule on the desirable correspondence of political instruments and goals). But this is another story of great actual importance.

Acknowledgments For critical remarks I am grateful to Prof. Fritz Helmedag, Dr. Karin Knottenbauer, Peggy Schindler and Dipl.-Ing. Stefan Wagner. See http://www.adolfwagner.eu

A Note on History

The history of circular flow analysis shows remarkably long periods of fashion and forgetting, of booms and busts. The banking and finance crises of our days should lead to a revival in modern macroeconomic flow analysis. Therefore, I add a list of relevant scientists of the past at this point. Circulation theory begins with Francois Quesnay (1694–1774) at Versailles. His leading question concerned the agricultural circulation within a year: advances of the landowners and sustainability of the circulation year by year. The outcome was the famous zigzag in 1758 and accounting-representations in 1766. Rarely mentioned is Joseph Lang (1775–1837) at Freiburg with publications of 1807 and 1811 (see Uebe 1988). Well-known, however, is Karl Marx (1818–1883) with his "tableau économique" regarding the reproduction of capital in societies (transactors: capitalists, workers, consumption goods sector, investment goods sector). Besides Quesnay, Lang and Marx, we have to consider Eugen von Böhm-Bawerck (1851–1914) and Hans Peter (1898–1959). Hans Peter (see Gilibert 2008, p. 794) founded the Tuebingen tradition of circular flow analysis. The Tuebingen school includes Hans Peter (1889–1959), Erich Preiser (1900–1967), Carl Föhl (1901–1973), Wolfgang Stützel (1925–1987), Helmut Reichardt (1922–2009), Ernst Helmstädter (geb. 1924) and Alfred E. Ott (1929–1994).

Literature

Allen RGD (1965) Mathematical economics, 2nd edn. Macmillan, London

Allen RGD (1968) Macro-economic theory. A mathematical treatment (first printing 1967). Macmillan, London

Bundesministerium der Finanzen im Benehmen mit der Deutschen Bundesbank (2006) Verordnung über die Liquidität der Institute (Liquiditätsverordnung – LiqV) vom 14. 12. 2006, Berlin

Gilibert G (2008) Circular flow. In: Durlauf SN, Blume LE (eds) The New Palgrave. Dictionary of economics, vol 1, 2nd edn. Palgrave Macmillan, London, pp 790–794

Helmstädter E (1981) Kreislauf und Reproduktion. In: Mückl WJ, Ott AE (eds) Wirtschaftstheorie und Wirtschaftspolitik. Gedenkschrift für Erich Preiser. Passau, Passavia Universitatsverlag, pp 145–157

Ihring F (1998) Die Tübinger Konzeption des Wirtschaftskreislaufs. Basel, Tübingen

Kaldor N, Trevithick J (1981) Geldtheorie und Geldpolitik V: aus keynesianischer Sicht. In: Handwörterbuch der Wirtschaftswissenschaft, vol. 3. Stuttgart u. a. O., Gustav Fischer, pp 412–423

Krelle W (1978) Die kapitaltheoretische Kontroverse. Test zum Reswitching-Problem. In: Zeitschrift für Wirtschafts- und Sozialwissenschaften, vol. 98, pp 1–31

Krugman P, Wells R (2006) Economics, New York; [German (2010): Volkswirtschaftslehre, Stuttgart.]

Mankiw NG, Taylor MP (2006) Economics, London; [German (2008): Grundzüge der Volkswirtschaftslehre, 4. Aufl., Stuttgart.]

Metcalfe J, Steedman I (1991) Mr. Harrod and the classics. In: Rima I (Hrsg) The Joan Robinson Legacy, London, pp 61–79

Peter H (1954) Mathematische Strukturlehre des Wirtschaftskreislaufs. Göttingen, Schwartz

Peter H, Reichardt H (1959) Kreislauf, ökonomischer, Artikel in Handwörterbuch der Sozialwissenschaften (HdSW), vol. 6, pp 348–357

Pokropp F. (1972) Aggregation von Produktionsfunktionen. Klein-Nataf-Aggregation ohne Annahmen über Differenzierbarkeit und Stetigkeit. Berlin u. a. O, Springer

Raabe KG (1967) Vorausschätzungen der allgemeinen Wirtschaftspolitik (ohne ökonometrische Modelle) am Beispiel der Bundesrepublik Deutschland. In: Allgemeines Statistisches Archiv, 51, Bd., pp 10–44

Raabe KH (1974) Gesamtwirtschaftliche Prognosen und Projektionen als Hilfsmittel der Wirtschaftspolitik in der Bundesrepublik Deutschland. In: Allgemeines Statistisches Archiv, 58, Bd., pp 1–31
Ramser HJ (1987) Beschäftigung und Konjunktur. Versuch einer Integration verschiedener Erklärungsansätze. Berlin u. a. O, Springer
Reichardt H (1967) Kreislaufaspekte in der Ökonomik. Basel, Tübingen
Richter R (1979) Makroökonomik, statisch. In: Selten R (ed) Handwörterbuch der Mathematischen Wirtschaftstheorie, Band 1. Wirtschaftstheorie, Wiesbaden, pp 173–189
Schlicht E (1990) Local aggregation in a dynamic setting. In: Zeitschrift für Nationalökonomie, vol. 51, pp 287–305
Schmidt H, Ketzel E, Prigge S (eds) (2001) Wolfgang Stützel. Moderne Konzepte für Finanzmärkte, Beschäftigung und Wirtschaftsverfassung, Tübingen
Seitz T (1989) Innovation, Marktdynamik und Preisbildung. In: Seitz T (ed) Wirtschaftliche Dynamik und technischer Wandel, Stuttgart u. a. O., pp 35–52
Statistisches Amt der Europäischen Gemeinschaften (1976) Methodologie der Input-Output-Tabellen für die Gemeinschaft 1970–1975, Sonderreihe, 1–1976. Luxembourg, Amt für Amtliche Veröffentlichungen der Europäischen Gemeinschaften
Stützel, W (1959) Ist die "Goldene Bankregel" eine geeignete Richtschnur für die Geschäftspolitik der Kreditinstitute? In: Deutscher Sparkassen- und Giroverband (ed) Vorträge für Sparkassenprüfer (Kiel, 8. – 11. 9. 1959), Stuttgart, pp 35–51
Stützel, W (1978) Volkswirtschaftliche Saldenmechanik. Ein Beitrag zur Geldtheorie, 2. Aufl., Tübingen
Uebe G (1988) Joseph Lang aus Freiburg. Eine Kreislauftheorie von 1807 zwischen Quesnay und Marx. In: Jahrbücher für Nationalökonomie und Statisti, vol. 204, pp 36–47
United Nations (UN), International Monetary Fund (IMF), Commission of the European Communities (EC), Organisation for Economic Co-operation and Development (OECD) (1993) System of National Accounts 1993. Brussels, New York, NY
Wagner A (1857) Beiträge zur Lehre von den Banken, Leipzig
Wagner A (1983) Angewandte Input-Output-Analyse in der Bundesrepublik Deutschland unter dem Aspekt ökonomischen Gleichgewichts. In: Enke H, Köhler W, Schulz H (eds) Struktur und Dynamik der Wirtschaft, Freiburg, pp 437–446
Wagner A (1989) Makroökonomisches Kreislauf- und/oder Marktgleichgewicht. In: Jahrbücher für Nationalökonomie und Statistik, vol. 206, pp 510–516
Wagner A (2003) Strukturwandel, Arbeitslosigkeit und Verteilung. Basel, Tübingen, pp 571–578
Wagner A (2009) Makroökonomik. Volkswirtschaftliche Strukturen II, 3 edn, Marburg

Money, Credit, Capital and the State

On the Evolution of Money and Institutions

Hardy Hanappi

Abstract This paper combines several important arguments, which have puzzled economic theory for decades, to arrive at a more adequate description of the current global crisis. The main theoretical innovation is to view the long-run economic evolution as a stepwise evolution of money forms. Moreover, as indicated in the title, this development of money forms is closely linked to the development of social institutions, in particular, state institutions. Capital, the most recent form of money, today has to be understood as an omnipresent algorithm, as a growth imperative implicit in social institutions and internalized models. The task of evolutionary political economy thus will be to provide an adequate theoretical counterpart to mirror these processes. This paper explores how far a careful reconsideration of received economic theory can contribute to this task.

1 Money

All human societies with a developed political economy are monetary economies. The very existence of the interplay of production and consumption activities of a society's members implies that periodically reappearing habits, of relations between activities, become visible and are memorized. At a certain point of development, the repetitions—using days and years as time units—are externalized as signs on physical carriers. Written language serves as a special device to adjust and to regulate activities according to processes going on in the natural environment of society. The immediate importance of the new device is evident: advantageous behavior can be transmitted across generations without relying on spoken language only. Sign systems on physical carrier systems thus acquire an important

H. Hanappi (✉)
Institute for Mathematical Methods in Economics, Vienna University of Technology, Vienna, Austria
e-mail: hanappi@econ.tuwien.ac.at

G. Buenstorf et al. (eds.), *The Two Sides of Innovation*, Economic Complexity and Evolution, DOI 10.1007/978-3-319-01496-8_13,

social role, a metaphysical (i.e. more than physical) status, which stabilizes social evolution.[1]

In economic theory, the three basic functions of money usually mentioned in standard macroeconomics—'store of value', 'unit of accounting', and 'medium of exchange'—all refer to this specific role.[2] As for growing societies, production is split up in time and space, as are its services and products. To bridge time and space, a memory of the performed activity is materialized as a sign on a physical carrier. This unity of physical matter and its relation to the members of society stores what is called 'social value'; its more mundane name is 'money'. Since contributions to social evolution are so different in kind—and even more so the more developed and complicated the society becomes—intra-social organization calls for the measurement of relative social value.[3] Again, the sign system of money can help by adding a scalar size, the units of social value accounting, to the material carrier. If these two prerequisites are given,

(1) the commitment of the members of society to accept a system of social value relations expressed on a physical carrier system;
(2) the acceptance of its specific quantitative expression as a set of certain relative money amounts;

Only then is exchange of social value via money as the medium of exchange possible. The existence of money thus coincides with the existence of social organization, already at a very early and primitive level.

But even at this lowest level, several implicit features of a monetary economy become visible, features usually not explicitly articulated in standard economic theory. Acceptance of a certain money system can be either voluntary or forced upon the members of society—or something in between. In any case, acceptance reflects the ***power relations*** within a society.[4] If power is concentrated in the hands of a small group of members of society, it is evident that the portrait of the leader of this small group on the coins used in this society should keep authority of the powerful alive. In more democratically organized systems with power only temporarily transferred to institutions, and recurrently checked by feedback mechanisms involving all members of society, the monetary system carries the marks of the

---

[1] There is, of course, an extensive body of literature discussing the historical and logical roots of sign systems such as money. European social scientists representing the academic turn of the New Left in the aftermath of the youth revolt of the late 1960s rediscovered this topic, e.g. (Müller 1977; Thomson 1972; De Brunhoff 1973).

[2] Compare (Mankiw 2010, Chap. 4).

[3] This interwoven character of the social value system and the institutional framework of social institutions is also reflected in modern mainstream economics, see (Walsh 2010), or (Mishkin and Eakins 2011).

[4] Some authors have emphasized the strong connections between the power structure implicit in money relations, and those existing in other social domains, e.g. symbolically amplified cultural hierarchy or gender hierarchy. In some cases, it is even suggested that the former is determined by the latter. Compare (Heinsohn 1984).

specific institutional setting prevailing in that society. The borders between voluntary and forced participation in the prevailing monetary system increasingly get blurred if one considers the next long-run trend.

In ancient Greece, the hegemony of Athena's coin, the famous silver owl, needed the rich silver mines close to the city as well as the mighty fleet of Athens. The threat of punishment in case of disobedience was very visible. This obvious backup for the power of Athena's monetary system gave it a very concrete character, a metallic taste. Since then, money has experienced a ***process of ever increasing abstraction***. With the establishment of more sophisticated exertions of power, banknotes could free the availability of the monetary carrier medium from the pressures of production of precious metal. Abstraction thus proved the primacy of the social sources of money's force, as compared to the inherent social value of the carrier medium. The social sources of voluntary subordination under a monetary system became manifold and more binding as the next best alternative—voluntary disobedience—was losing ground quickly when the growth of monetary economies took off.[5] Nowadays, money is just a rapid change of patterns on a computer screen, and so the mission of the abstraction process starting with the original problem has been completed. The overwhelming majority of human individuals have no choice but to voluntarily join monetary rules; on the way, the content of the concept 'voluntary' has lost its meaning.

Nevertheless—and despite the highly complicated setting of institutions and democratic feedback loops[6]—contemporary money still is a sign system that refers to abstract social value. For a single individual, one out of seven billion humans worldwide, the evaluation of relative social value of all the activities and products with which it is concerned certainly cannot be derived by insight into this complicated, global production structure. This difficulty, the inability to grasp the complexity of the global production, lies at the root of subjective value theory, of microeconomic theory as it has been taught since Walras, Menger and Jevons (around 1874). This theory thus is not just an ideological vehicle to fight the macroeconomic exploitation theories of the intellectuals flirting with the labor movement of the nineteenth century. Subjective value theory addresses a real problem of households: How to order purchases if a given amount of monetary income can be spent—the starting point for axiomatic utility theory.[7] But as this set of decisions falls apart from an understanding of production and reproduction of the overall political economy, the human individual is separated into an economic, utility maximizing entity and a political entity engaging in the multi-layered

[5] See (Davies 2002) as reference for a detailed history of this process.

[6] In a democracy, the institutional setting is repeatedly restructured by a system of second-order institutional processes (including electoral processes) designed to enable entry and exit of first-order institutions. This, of course, implies that a dangerously smooth transition to non-democratic governance can occur.

[7] A faint memory of the link to the overall social process can be seen in J.M. Keynes emphasis on the concept of effective demand. In denouncing needs, which cannot be translated in money terms as less important, he intuitively subscribes to classical political economy.

processes of more or less democratic feedback control. Production units, the other micro-unit covered by microeconomics, is the archetype of the decision-making entity put in the center of analysis. It is their (hypothesized) stylized decision problem that is grafted onto household behavior to describe it as utility maximization, the analogue to profit maximization. Again, entrepreneurs are schizophrenic: On the one hand, there is the input cost minimizing (and in the sequel, profit maximizing) entity that is completely ignorant as to the political evolution of its environment, and on the other hand, there is the legal institution with special status in the legal system and vested interest political evolution.[8] For both microeconomic types of agent, money (i.e. social value) is the exogenous constraint that makes needs and profit possibilities appear scarce. The arrangement of social activities and their guiding institutions seems to be out of range of the microeconomic discourse, but if money is interpreted as a first form of a regulating device shaping the evolution of the former, then the old tenets of classical political economy appear on the theoretical horizon again. The sign system of money still refers to the arrangement of human activities, to the setup of time spent at an enormous variety of occupations worldwide. To see how a crisis of social evolution is translated into, and amplified by, a crisis of money forms, the renaissance of political economy in the form of Keynesian macroeconomics has to be briefly reviewed.

I will leave this argument on the role of money at this point, and take it up as a loose end in the last section.

2 Credit

At first glance, credit seems to be just another face of money: Accepting a coin instead of a certain amount of a commodity—usually in exchanges at markets—can be understood as giving credit, literally to believe (credo) in the validity of a sign on a carrier medium. The point, of course, is that it is not the partner in the exchange transaction to whom this believing refers. It rather is belief in the enduring trustworthiness of the social system which provides the environment for the exchange act. Money interpreted as credit enlarges its original function of bridging space and time for a species that has developed a common consciousness of mutual commitment. When, towards the end of the Middle Ages, merchants in Venice received credit from the wealthy local banks, this credit always came in the form of money.

But not all money can be considered to be credit in the more narrow sense introduced by these early global merchants and explorers. To some extent, the distinction is merely based on the temporal scale that is relevant. The coin in the pocket is only short-run property, easily exchanged at the next opportunity, whereas

[8] The classics saw this dichotomy very clearly and dubbed it the double existence of the citoyen (the politically emancipated member of society) and the bourgeois (the owner of a factory engaged in profit maximization).

money taken as credit by the above mentioned merchant was used up during a journey that often took years. But this notion of difference in physical time is just the appearance of a more substantial difference, based on the emergence of an economic time scale. The coin in the pocket remains inactive; its sole function is to preserve a certain amount of social value by freezing it as a sign on a piece of metal. The money given as credit to the merchant not only remains property for a longer physical time span, but during the journey it indeed changes its form as well as its social value. By the actions of the merchant, a new social archetype, money as credit, becomes a process. This process, due to the ongoing change of money and commodity forms representing different systems of social value in different countries, produces its own type of time: economic time. Economic time runs parallel to but not synchronous with physical time.[9] Consider two journeys of merchants, the first taking 3 years of physical time, the second taking 6 years. Assume that at the end of the first journey, when the merchant comes back to Venice and sells all the commodities he has brought, that he owns an amount of money, which is five times as high as the sum with which he started. Assume further that, after the second journey, the analogue amount of money owned by the merchant is ten times the original credit. In that case, the speed of economic time of the two journeys would be equal, since the (fictitious) growth rate of credit-money per physical time unit is the same. It is the growth rate of the amount of credit-money which serves as measure the economic activity; if it falls to zero, then credit-money collapses and remains just money.[10]

While this view of credit emphasizes its role in stimulating the economic activity of merchants, it nevertheless is linked to a specific perspective on its relation to savings. For the classical political economists, the progressive role of those who use credit-money to transform it into economic activity was evident. Consumption was basically understood as the consumption of the feudal class, and the heroes of the new economic era were economic agents, who—contrary to the consuming feudal parasites—transformed money into economic growth.[11] Credit for the purpose of consumption was thus of no economic significance; if rich merchant families

[9] Economic time did exist before the emergence of credit, but it remained invisible since it was hiding behind the rhythm dictated by the seasons, which was synchronous to the major field of economic activity: agriculture. The treatment of the concept of time in political economy is remarkably underdeveloped, which mainly seems to be due to a somewhat blind adoption of the formal apparatus of the natural sciences. Indeed, some insights from sociology, e.g. (Elias 1993), still wait for to be appropriately appreciated in political economy.

[10] In modern economic jargon, this is often expressed as ‚money being a security with an interest rate of zero’. Evidently the determination of the set of securities bearing an interest rate of zero is further complicated if inflation rates (and the different possibilities to calculate them) are taken into account.

[11] From Adam Smith’s arguments for the source of wealth to Keynes revival of the idea of the ‘euthanasia of the rentier’, progress has been identified as the advance from feudal (over-) consumption to productivity-increasing re-investment of entrepreneurs. This common anti-feudal trait, of course, is in stark contrast to the basic differences of the worldview of these two godfathers of modern economic thought. Compare (Foley 2006).

provided money for ‘Il Principe’, then no economic growth process was expected.[12] On the other hand, the working class was considered to be involved in the process of accumulation only as a passive element, using money as a medium of exchange for immediate physical reproduction, and never as credit-money. Credit for consumption thus was thought to be insignificant in explaining the essence of credit-money. Savings of non-feudal households played either the same role as inventories for production units— a tool to smooth the stream of income — or, if they became more systematically growing as in the rich city-states of northern Italy, they were collected in banks financing merchants. This latter process thus clearly represents emergence of a new division of labor within the class of non-feudal rich families[13]: one group provides credit-money, while the other uses it for exploration and trade. Note that this emergence needed at least two elements, (i) a certain vacuum within feudal power structures and (ii) a certain level of money hoardings.[14] It might well be that there is a more general—and acute—lesson to be learned for the emergence of institutions in the current situation.

Only when the labor movement gained significance towards the end of the nineteenth century did the savings of worker households start to play a macroeconomic role. Again the concurrent ideological turn to microeconomic reasoning prevented mainstream economic theory from recognizing the importance of this development.[15] It needed the Great Depression and, in the sequel, Keynes, Schumpeter and their followers, to reshape classical political economic theory to grasp some of the elements of the new era. Keynes rediscovered the importance of circular macroeconomic money flows, while Schumpeter—drawing to a considerable extent on Marx’s ideas—highlighted and sharpened the implications of the historical mission of entrepreneurs and innovation.[16] But by the time that happened, the form of credit-money had already developed into a new dominating process, capital. Credit-money as money for consumption thus only appears when the next metamorphosis of money forms has taken place.

Return now to the original add-on, which makes credit-money a historical and logical bridge between money and capital. While the existence of money is just a

---

[12] What actually was expected from the feudal sovereign was a guarantee for political stability. Compare (Machiavelli, 1988 (1532)).

[13] At this point, the concept of a social class enters the argument. It turned out (and still turns out) to be of central importance for an analysis of capitalism. Despite the renaissance of meso-economics challenging the bipolar world of micro- and macroeconomics, (global) class analysis still has not reached the theoretical status it should have. Compare (Wright 2005).

[14] The first globally hegemonic country of merchant capitalism, the Netherlands, is another good example of the importance of these preconditions.

[15] In this context, the work of Veblen—certainly a maverick economist in his time—is remarkable, since he directed attention to the consumption behavior of the non-workers, the ‘leisure class’ [see (Veblen 1899)]. Deriving class membership from the ‘conspicuous consumption’ habits of persons rather than from their roles in the production process, he implicitly observes that to get credit for consumption purposes is a defining characteristic of the leisure class.

[16] See (Catephores 1994) for a more detailed explanation of this view.

reflection of the generally recognized unity of a prevailing social setting, the emergence of credit-money is a partial negation of money's universality: With the credit given by a specific member of society to a specific other member, the concerned money amount is not simply secured by the state monetary authority. It is additionally secured by a private contract between the two agents involved, a contract that itself is correctly called a security. Note that with this new development of a specific kind of credit-money there does emerge the concept of private economic agents—as opposed to the physical individuals inhabiting the world of simple commodity producing societies.[17] Note further that the use of contracts implies the emergence of a corresponding ***specific law system,***[18] which in turn is built up by ***a host of emerging institutions***. These institutions become necessary to assure that the procedures agreed upon by the involved parties in advance (as content of the contract) are actually executed as economic time proceeds, i.e. that commitment in the credit market becomes feasible.

The evolution of new institutions for new private economic agents evidently coincides with the use of credit-money—and this, of course, challenges the intellectual commentators of the time anticipating a clash with the already existing political institution of feudalism. When Montesquieu designed his famous idea of a division of power within the modern state, he did so on the basis of a careful comparison of empirically observed systems in different times and countries (Montesquieu 1748). Credit-money as a bridge to capital, i.e. the next form of money, also paves the way to an understanding of the form of political organization accompanying capital: the ***nation state***. This newly emerging political organization freed itself successively from its feudal bonds to provide an ***adequate structure of power relations*** for capitalism. It did so by monopolizing coercive power and institutionalizing the links between private economic agents. Again this argument will be taken up in the concluding section.

Finally, one side-effect of the credit mechanism used by merchants has to be highlighted, since it unconsciously prepared the next step of social evolution: With their successful trading activities, merchants were indeed starting to increase global, average labor productivity. By buying, transporting and selling commodities (and sometimes slaves) to increase their working credit-money, some pre-existing specialization in the different parts of the world they explored[19]

[17] It is remarkable how mainstream microeconomics from 1870 onwards (starting with Walras 1874; Jevons 1871; Menger 1871) systematically confuses private economic agents and biological individuals. The major reason for that deficiency is the complete neglect of the evolution of money forms.

[18] Laws thus are man-made and not innate economic properties of human individuals. Note the sharp contrast of this perspective of ***explaining the emergence and evolution of laws rather than discovering them***. Even the approach of experimental economics falls prey to this misconception of microeconomic ideology when it simply tries to discover innate economic laws differing from the ones stipulated by neoclassical doctrine, e.g. altruism.

[19] Compare, for example, the vivid description of global trade triangles in (Frank 1978).

entered global consumption, opened up new utility dimensions[20] or reduced average necessary labor inputs. Of course, such global effects were not recognized by the merchants themselves, and were additionally obscured by the fact that most advances were absorbed by the still powerful feudal elite. Nevertheless, ex post, an increase in global technological abilities as well as a widening of consumption spaces due to merchant (credit financed) activities is evident. With the next step of monetary evolution, this aspect of accommodating technological evolution proved to be one of the dominant elements of social progress.

3 Capital

If money not only is used as credit but assumes the form of a generally applicable program of accumulation, then it is called capital. Capital has all the features of credit-money but additionally, in the form of a mandatory algorithm, subordinates all strata of social organization. Credit-capital had explored and conquered society's environment. Capital also turns inside the more and more global society and its program takes hold of all humans and institutions to transform them into economic agents, into drivers of its abstract algorithm. Still, the appearance of the money involved in this process has not changed all of the properties related to the lower forms (general acceptance, importance of contracts and private economic agents, etc.), which still apply. There just is the metamorphosis into a general principle guiding the carrier systems instead of being guided by the latter.[21]

The program of the capital process in its most general form is rather simple and consists of the following commands: ***Capital Algorithm***

For each member of the set of currently possible visions do ('vision loop')

Produce a vision of specific entrepreneurial activity
Check expected wage cost
Check expected interest on credit-money (vulgo 'capital cost')
Check expected effective demand
Compute expected growth rate of capital
Estimate the probability to achieve that growth rate

(continued)

[20] The change in the dimension of the utility space—some new commodities enter, others vanish—is one of the most important blind spots of mainstream economics, which evolutionary economics promises to shed light on.

[21] Karl Marx had anticipated this, and chose 'Das Kapital' and not 'Die Kapitalisten' as title for his opus magnus. In the second preface to this book, he explains that the notion of 'capitalist' is to be understood as an abstract algorithm (a 'Charaktermaske'), and not as an immediate reference to physical individuals (Marx 1857).

(continued)

End of vision loop
Choose the vision yielding the highest utility of a mean-variance utility function
Check if the selected vision's utility exceeds the expected utility of a supplier of credit-money
If the lender's utility is higher, then perform the chosen project, else become a supplir of credit-money.

This innocent algorithmic prescription generalizes what merchants and their bankers did with credit-money by the end of the thirteenth century.[22] But what makes the difference is the fact that, in the course of the historic development of capitalism from merchant capitalism to industrial capitalism,[23] the abstract form of this algorithm proved to be universally applicable to all kinds of activities of economic agents. A look around contemporary OECD countries reveals the fact that there is almost no aspect of life that is not permeated by the workings of the capitalist program.[24] In a sense, the monetary core of the activities of certain groups in early merchant capitalism[25] has turned from outside trade to all types of inside activity. In the end—in (Hanappi 1989) this stage is called 'integrated capitalism'—not only production units but every household and every institution has become a private economic agent following the abstract algorithm of capital accumulation. Concepts such as human capital and competence capital show that the higher degree of abstraction that money gave rise to enables and opens up an incredibly wide field of possible application. It is thus not surprising that, in the history of economic thought, a sharp turnaround took place: ***the mirror image of the real course of economic development observed in its contemporary state started to be taken as its actual origin***. Transplanted into physical human individuals, from Robinson Crusoe to the more abstract homo oeconomicus, the private economic agent was considered to be the atom of 'social physics'.[26] From that

[22] Venice and Genoa had started to mint their own coins to support their conquest of world trade [compare (Braudel 1986, pp. 111–116)]. The less abstract forms of money are thus not simply substituted by a new form. Rather, they are only adjusted to accommodate the new hegemonic form of money. This could also be a lesson to be learned for the current crisis.

[23] For a detailed discussion of the stages of capitalism, compare (Hanappi 1989).

[24] The somewhat forgotten German social scientist Alfred Sohn-Rethel introduced an interesting hypothesis: Even the logical structure of humans' mental models is framed by the evolution of commodity producing societies (Sohn-Rethel 1978, pp. 103–133).

[25] Distinction by function separates bankers from merchants, distinction by location [following (Braudel 1986)] separates Brügge, Hanse cities, Northern Italian cities, Champagne, Antwerpen—and later Amsterdam.

[26] This expression was used by Auguste Comte to make clear that his vision of a future social science follows the example of the natural sciences (Comte, 1979 (1844)).

perspective, the true state of nature had been obscured in the past and only in full-fledged capitalism could the true and final character of social relations reveal itself. Once this final state is reached, history has ended—only some safeguarding to prevent external disturbances (modern economics calls them ‘shocks’) is needed.

It is interesting to see that with such a radical conceptual turn—mistaking a frozen mirror image as a parable of origin—not only evolutionary political economy becomes impossible; also money in its highest form of abstraction vanishes, since it becomes an innate feature of private economic agents. The current indecisiveness of mainstream economic advisers dramatically shows the impasse, which was taken a long time ago.

But the methodological turn of economic theory towards the crude atomistic perspective had several other severe consequences, too. Since the formalism adopted originally was a description of energy transformations of non-living, smallest elements of matter,[27] any description including the build-up of structures and clusters is simply impossible! The final issue emerging from that formalism in the natural sciences is the second law of thermodynamics, which states that, in the long-run, the stochastic trend towards an increase in entropy will prevail, i.e. a certain equilibrium state of (computable) maximum entropy will be approximated. The re-interpretation in microeconomic terms postulates this process as the working of market forces relating the owners of (‘scarce’) resources and in the long-run leading to a vector of relative prices—the correlate of maximum entropy in general equilibrium theory. Note that ‘prices’ in this context are exchange relations of quantities of commodities[28] and ***not*** a monetary expression of the social value of a unit of a certain commodity. To bridge—or, more ideologically interpreted, to disguise—this strange role of prices, the theory had to be complemented by the adoption of the so-called quantity theory of money. If maximum entropy is reached, vulgo ‘in general equilibrium’, the vector of relative exchange quantities can be translated into a vector of money prices by simply assuming proportionality to the amount of money signs (on carrier systems) in circulation as well as to an exogenously assumed speed of circulation. Evidently the money form used in the quantity theory of money is ***not*** credit-money or capital. In the pure form of general equilibrium theory (GET) ***there is thus no endogenously developed theory of***

[27] Compare (Smith and Foley 2002) for a detailed treatment of that isomorphism.

[28] It is surprising to see how, in 1871, one of the founding fathers of GET, Stanley Jevons, already spelled out its methodological break: “I have attempted to treat Economy as a calculus of pleasure and pain, and have sketched out, almost irrespective of previous opinions, the form which the science, as it seems to me, must ultimately take. I have long thought that as it deals throughout with quantities, it must be a mathematical science in matter if not in language. … The Theory of Economy thus treated presents a close analogy to the science of Statistical Mechanics, and the Laws of Exchange are found to resemble the Laws of Equilibrium of a lever as determined by the principle of virtual velocities. The nature of Wealth and Value is explained by the consideration of indefinitely small amounts of pleasure and pain, just as the Theory of Statics is made to rest upon the equality of indefinitely small amounts of energy.” (Jevons 1871, p. viii). Pleasure and pain are inborn features of a material smallest entity, just as properties of atoms in physics.

money. Money and prices are just as a veil thrown by a monetary authority over true and 'natural' exchange ratios.

A further dramatic consequence follows: If there is no theory of credit and capital, if the theory of social values is dissolved into predetermined preferences of a set of biological atoms moving towards its natural equilibrium via markets—then ***there is no room for an understanding of the growth of structure***, of exactly those processes that constitute the emergence of life forms (of all evolutionary forms) living as temporary contradictions to the law of entropy.[29] The neglect of the evolution of money forms therefore is just the tip of the iceberg of the methodological sins of the atomistic turn. Growth, the central concept around which all types of biological theories are built, is explicitly excluded by assumption.

And as a final, but pivotal, side-effect, growth in life forms tends to produce ***species*** and at the same time the ***exploitative relations*** linking these species.[30] While the rate of growth of one species might be advanced by slowing down growth of the exploited species, feedback enhancing the growth of both eventually is possible. In other words, exploitation is a dynamic concept measurable over well-defined historic time spans.[31] Recaptured from this perspective, the most advanced form of money, the universally applicable algorithm of capital, is just the abstract claim of social value to grow. The ways that such a growth of social value can take are not limited to the development of new forms of exploitation across and within species.[32] To explore in which sense omnipresent capital processes can be superseded, which features might survive and which elements will have to be replaced rather rapidly, goes a bit beyond the scope of this paper—though the concluding section will come back to this issue.

With all these deficiencies, the new microeconomic view at the turn of the century—though an apt ideological vehicle to attract a considerable part of the intellectual elite—was unable to grasp the two most important processes going on in actual economic development: ***technical progress and institutional evolution***.

To understand the former, it would have been necessary to treat capitalist production units as exploitation maximizing enterprises, and not just traders of the resources they own, forced to low prices by competitive markets. With the same

[29] While the Second Law certainly has the aspiration of an eternal truth, the stochastic character of that truth implies that temporary counter-movements can occur. This in turn implies that such a build-up of neg-entropy has a beginning and an end—carriers of life necessarily are born at a point in time and die after a finite amount of physical time.

[30] As with the simultaneous emergence of bankers and merchants on the one hand, and their relationship (credit-money) on the other hand, emergence of different species and their relations to each other are one and the same process.

[31] For a more detailed treatment of this idea, compare (Hanappi 2006).

[32] For the human species forms of exploitation of nature are usually combined with forms of the exploitation of one class of society by another class of society, of man by man. Since even in biology borderlines between species are hard to determine genetically, exploitative social relations play a central role for the structure between and within species. But note that exploitation, and thus class, is a dynamic concept, subjected to evolution.

argument, it would have become clear that the emergence of new institutions—of the bourgeois class as well as of the newly emerging labor class[33]—was enforced by the contradictions between the different groups, the different classes.[34] The endogenous emergence of institutions stays out of reach for any theory that insists on the notions of the 'representative household' and the 'representative firm'. But both developments—technical advance as well as institutional evolution—actually were extremely strong and had the profoundest feedback influence on the path of economic evolution.

Inadequate theory is bound to be faltering—at least in the long-run. When the Great Depression proved the assumption of the effortless final arrival at the paradise of free market interaction to be definitely wrong, critics of the received doctrine—which itself necessarily remained mute—had their say.[35] Schumpeter attacked the common wisdom of the economist profession by substituting the equilibrium of traders of resources by a diversity of active entrepreneurs eager to push markets out of equilibrium.[36] Keynes added the political institution of the nation state and an independent influence of money oriented behavioral traits, e.g. an independent investment function, to replace the scanty models of his teacher, Alfred Marshall.[37] Rudolf Hilferding, creatively extending some of Marx's insights, went even further and tried to incorporate the latest development of capitalism—he insisted that, on the way towards more oligopolistic market structures, a new form of capital was emerging: finance capital.[38] The only Austrian Nobel prize laureate, Friedrich Hayek, critically and lucidly remarked: '*What I complain of is not that this theory [the quantity theory of money] in its various forms has unduly usurped the central place in monetary theory, but that the point of view from which it springs is a positive hindrance to further progress. Not the least harmful effect of this particular theory is the present isolation of the theory of money from the main body of general economic theory*.' (Hayek 1931, p. 4).

These and other criticisms lead to a revival of ideas closer to classical political economy, and as a new departure within economic theory, it came with a new brand name: macroeconomics. In hindsight, it seems to be rather obvious that it never really developed into a common, generally accepted view synthesizing all essential aspects of twentieth century capitalism.[39] Despite the amazing theoretical progress in many specialized areas, no coherent set of theories able to describe the essential

[33] The strongest growth of labor union membership occurred just in the three decades before World War One.

[34] Institutions, therefore, typically can either be a vehicle serving as a focal point for one of the involved groups, or constituted as a (temporary) freeze of a compromise reached.

[35] In this respect, the current crisis shows the same characteristic.

[36] See e.g. (Schumpeter 1939).

[37] See (Keynes 1936).

[38] Compare (Hilferding 1910).

[39] Paul Samuelson's vision (still a pupil of Schumpeter) of a ‚neoclassical synthesis'remained a fragment in that respect—and later collapsed completely when the so-called 'microfoundation of macroeconomics' failed.

characteristics of the development of the world economy emerged. One reason might have been the extremely shaky course global political economy took in the twentieth century.[40] An overarching theoretical construct would have needed much more effort with respect to the two formerly mentioned blind spots of mainstream economic theory: technological progress and institutional evolution.

Macroeconomics as part of economics was firmly established and culminated in a widely accepted formalization of Keynes' central ideas provided by John Hicks: the IS-LM framework. There, the idea of the importance of state intervention degenerated to shifts of both schedules (IS and LM) in an output-interest rate diagram due to government action (fiscal or monetary policy). The idea that money processes should be intrinsically included in any model of a monetary economy was reduced to two theoretical innovations: (1) A money demand function, which not only included the traditional transaction cash motive but also demand for 'speculative purposes'; (2) an independent investment demand, which compared an expected internal rate of return with a prevailing market interest rate. Both arguments refer to the process of using credit-money to achieve growth and thus comply with the stage of capital.

An additional feature of Keynes' model—perhaps its most important property[41]—was that it revived an old idea going back to the school of Physiocrats in the eighteenth century.[42] In each year, within a closed geographical region, the total amount of money is bound to stay constant, but has to follow a certain circular flow mirroring the needs of different social classes in the course of the year. When in agricultural societies seed and harvest set the rhythm, the year was a natural beat for the whole economy. Keynes, as long as he looked only at flows, could suggest a similar scheme: If new systematic build-up of inventories occurs, then total output during a year has to be identical with total demand (both in money terms), and since the different uses that this demand is channeled to can be neatly structured, there emerges an additional modeling constraint.

It is remarkable how this simple amendment that linked aggregate supply and aggregate demand (both in monetary terms), together with Schumpeter's suggestion of disequilibrium dynamics, was able to stimulate a first wave of non-linear macrodynamics.[43] But unfortunately, Keynes' attitude to state behavioral rules

[40] There are good reasons why the eminent historian Eric Hobsbawm has dubbed this century the "Age of Extremes", see (Hobsbawm 1996).

[41] One of the most important side-effects was the establishment of statistical offices, which in most advanced countries started to collect data along the lines of Keynes' circular flow variables. Economically relevant relationships were suddenly assumed to be found in the data collected rather than as innate properties of human brains.

[42] The leading figure of this school producing the famous Tableau éonomique was Francois Quesnay (Quesnay 1758).

[43] As Paul Krugman later correctly noted, Richard Goodwin, (Goodwin 1955)—another pupil of Schumpeter -was the champion of that movement. See (Krugman 1996, p. 63). Goodwin in his later work explicitly hinted at his intention to combine Marx's ideas with Keynesian modeling and insights of Schumpeter to contribute to what he called the MKS-tradition.

always with variables in real money terms, i.e. postulating that economic agents are always fully aware of inflation, prevented the early model-builders from taking seriously Schumpeter's warning that a well-developed banking system is pivotal for innovation. Macrodynamics of the 1940s and 1950s became a disappearing fashion.

Nevertheless, a more adequate picture of what essentially was happening in the long-run of capitalist development was in the air. Schumpeter, inspired by Nikolai Kondratiev,[44] proposed to single out innovations—and in the sequel, their drivers, the economic agents he called entrepreneurs—as the central elements of capitalist progress. The mechanics proposed by Schumpeter and Keynes in principle were quite clear and not too far away from the actual working of the system. It may be sketched as follows:

In repeated cycles, households save part of the money they earn, transfer it to banks, which in turn provide credit-money for entrepreneurs. Competitive markets force entrepreneurs not only to invest, but also to increase the labor productivity of existing production processes[45] and to introduce new products and services. The increased output emerging that way could either be added to the stocks of exploited profit in banks (hence the banks' central role in searching for promising entrepreneurs) or could be given to the ever more organized labor class to secure political stability as well as effective demand. Since these rather sophisticated dynamics surely look unstable, the capitalist state is necessarily intervening—either politically (Schumpeter's view) or economically (Keynes' view). This vision of the process flirts with Hegel's "List der Vernunft" as well as with Mandeville's "private vices to public virtues".[46] As entrepreneurs strive for maximum profit, the structural constraint of competition forces them to do something beneficial for the whole society; specifically, they drive innovation—and they usually are not aware of that fact. Above all, this view seems to be close to what Karl Marx seems to have envisaged as the historical mission of capitalism in his manifesto (Marx 1848).

Though not adequately formalized yet, the success of the system described in the Keynes-Schumpeter perspective would clearly be supported by empirical observation. The social contract implicit in this working of twentieth century advanced countries was extremely successful in increasing GDP per capita. Diagram 1 shows the data for 12 Western European countries[47] first collected by Angus Maddison (Maddison 2006). This shows the explosion of technical progress, which has to be explained by economic theory. Note that GDP is measured in real terms and that per capita GDP is based on total population, not on employment. Labor productivity would be based on employment and thus would show the more production process oriented development, whereas the line shown in Diagram 1 concerns the average welfare effect of technological advance.

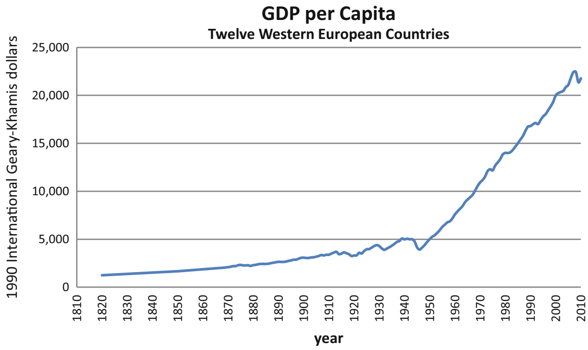
[44] Compare (Kondratiev 1926).

[45] A particularly interesting and recent empirical study of this link is (Ilyina and Samaniego 2009).

[46] See (Hegel 1807) and (Mandeville 1714).

[47] They originally were chosen due to data availability: Austria, Belgium, Denmark, Finland, France, Germany, Italy, Netherlands, Norway, Sweden, Switzerland, and the United Kingdom.

Diagram 1 Growth of GDP per Capita



But Keynes' reintroduction of the cyclical character of economic activity, despite its merits, had crucial, even devastating short-comings. Instead of modeling accumulation and exploitation, his concentration on circular flows falls back on the counterfactual assumption of equilibrium. Keynes' *bon mot* that 'in the long-run we are all dead' hides the fact that his basic setup necessarily is restricted to short-run considerations—without any references to empirically observed analogies. As soon as medium-term developments appear, this setup collapses. The simplest case was taken care of by early growth models, e.g. the first one constructed by Keynes' colleague Roy Harrod.[48] The simple extension mainly concerned a definition: Capital is defined as the stock of accumulated net investment.

From the point of view of evolution of money forms argued so far, this is a serious misconception. While the amount of money signs in a given area certainly can be measured at well specified points in time, and an increase of that amount over time might be computed, this still remains measurement of money in its first form and cannot explain the role of credit and capital. Even nineteenth century classical political economy knew better when it coined the phrase 'madame la terre et monsieur le capital'. Land clearly was seen to be the passive factor of production; possible output was usually thought to be a little bit less than proportional to its extension. This is a far cry from the confusion occurring later when land—now called 'capital'—was assumed to have decreasing marginal returns. On the other hand, the active role of the classical 'monsieur le capital' vanished from the production function until Schumpeter let him in again as magically shifting functions due to his 'entrepreneurial spirits'. One of the most exciting tasks of evolutionary economics that still remains is to put more flesh on Schumpeter's sometimes rather naïve view of this innovation process. Above all, little has been achieved to further our understanding of capital as a process of the most developed ***monetary*** form. How important that is can again be grasped by a glance at Diagram 2 showing the long-run evolution of the Dow Jones index.

Since this index provides an evaluation of the most important corporations in the world performing the capital program, it reflects how successful capital works. Noteworthy is the great take-off that appears after 1982, and not immediately after

[48] Compare (Harrod 1939).

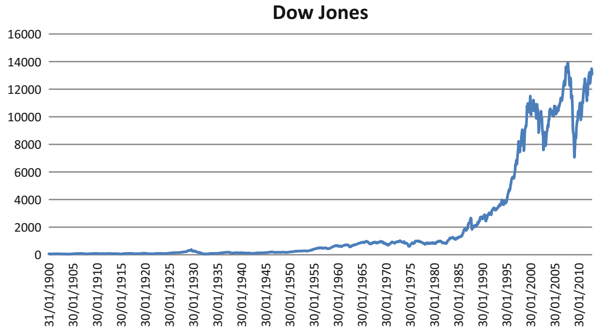


Diagram 2 Dow Jones Index as a description of the success of capital

World War Two as in the previous Diagram 1. This highlights the two phases of accelerated accumulation: one immediately after the war, basically enhancing welfare, and a second phase from 1982 on, which left welfare on its old track but added substantial accumulation to the Dow Jones measure.[49] The economic interpretation of the two phases is straight forward: While Western Europe over the whole after-war-period until 2007 experienced a strong increase in economic power (power of population to produce GDP) and general welfare (GDP distributed to citizens)—everything on average—there is an additional upward push in the evaluation of the globally most relevant corporations since the beginning of the 80s. This second phase thus evidently coincides with the broad political roll-back to conservative economic policies in most OECD countries, Reagan in the USA, Thatcher in the UK, Kohl in Germany, and others.[50] Seen from the international perspective, this political switch, after a long-period of a catch-up race of Europe and Japan, heralded the somewhat surprising second wave of even stronger unchallenged US hegemony, culminating in the break-down of the Soviet Union in the early 90-ties.

Obviously, the working of capital on a global scale became even more strongly linked to political evolutions in this second phase than before. In the vision of most of mainstream economists even today, the discipline of economics concerns specific mechanisms (mostly market mechanisms) working in a vacuum of direct coercive power. All such power is thought to be monopolized in an anonymously governing political entity securing economic rules—and the study of this entity falls out of economists' concern since it is covered by political science. As argued above, this political entity never has been anonymous, but rather resides at the core of economics, providing its indispensable monetary authority. This monetary authority mainly comes in two forms: A guarantee of the ***validity of the sign system***

[49] The DJI (industrial average) is the most widely accepted measure of capital activity with the longest historical time series.

[50] In a recent paper, this second phase also is characterized as a worldwide inflow of capital to US capital markets. Compare (Mendoza et al. 2009). It is accompanied by a relative shift of the US portfolio (as compared to other countries) towards more risky capital algorithms. Probably the stronger military stance of US policy enabled this shift.

representing social value, and the ***provision and adaption of the rules of the economic games.*** In particular, the second element neatly fits with the definition of capital put forward in this paper: It is the ensemble of the change of rules in nation states, the reframing of the program environment of the capital algorithms, which is responsible for the boom after 1982.

In other words, the long-run enhancement in labor productivity became superimposed by institutional evolutions initiating a new spurt of capital increasingly centered in the USA. Institutional settings became incredibly important for capital[51]; in short, an old player in classical political economy reentered the center of the stage: the state.

4 The State

Nation states in their new, non-feudal form are the institutional correlates to the evolution from credit-money to capital. Though in most countries the feudal class officially remained in power until World War One, the underground erosion of its influence started much earlier. When its rule finally broke, the world slipped into the deep troubles of the twentieth century,[52] which in retrospect can be understood as a search for new institutional solutions. The two remaining regimes, which—contrary to the fascist model—survived after WW2, were the new integrated capitalist state and the Stalinist production system.[53]

The development of institutional settings, as already mentioned in the previous section, is the second blind spot of mainstream economics today. It has received even less attention than technical progress. Of course, there is a fast growing literature discussing and describing the diversity of contemporary capitalism,[54] and even the implications of cultural and institutional diversity on decisions of transnational corporations have been extensively empirically studied.[55]

---

[51] A good example is telecom firms for which the national regulations of a country are the most important influence on profits. More generally, transnational corporations' success critically hinges on using different national regulations for labor market regimes and tax regimes.

[52] A second look at Diagram 1 reveals that these troubles also materialized as welfare losses: World War One, the Great Depression, and World War Two.

[53] Compare (Hanappi 1994, pp. 103–162) for a more detailed description of the latter.

[54] An interesting comparative study of five types of contemporary capitalism was presented by Bruno Amable (Amable 2003). For a survey of issues related to that variety, see (Elsner and Hanappi 2008).

[55] Some of the most promising contributions come from a group of Dutch economists, who originally tried to nail down the theoretical framework of New Economic Geography to answer empirical questions of spatial economic policy. See (Brakman et al. 2007, pp. 267–405). Another good example of empirically interesting conclusions with respect to firm evolution is (Lechevalier 2007).

Nevertheless ***the evolution of this diversity*** has not been conceptually particularly convincing. There has been some effort to collect different perspectives from heterodox economics, see e.g. (Hodgson et al. 2001), but a synthesizing approach still seems to be out of reach.

It is not surprising that attempts to conceptual institutional evolution along the lines of Darwin's ideas on biological evolution of the human species emerged early on, with Herbert Spencer readily introducing the analogy between biological selection and survival in competitive economic markets.[56] A thorough discussion of the issues surrounding 'Social Darwinism' and the role it played for fascism would go beyond the scope of this paper, it is nevertheless evident that, after WW2, the misuse of Darwin's concepts through fascist propaganda let social scientists shy away from direct applications of biological concepts to social developments. In retrospect, not much has been lost by the neglect of these early developments. The evolution of institutional settings was not in the focus of social Darwinism anyway, Spencer's atomistic view (nineteenth century market liberalism) had been substituted by archaic, hierarchical systems with little or no explanatory force.

What actually happened in institutional structures since 1945 is, above all, ***a strong trend towards more continental political entities***, the emergence of continental units. Though there is considerable change in the list of nation states, too,[57] the social innovation of the last 60 years clearly has been 'continental units in a common global context'. The most remarkable institutional social innovations date back to the early years of that era: the World Bank, the IMF, WTO (former GATTS), and the UN. Note that one important international rule system did not survive the early 1970s and thus is a good candidate for the explanation of the following discontinuity in 1982: The Bretton Woods system of fixed exchange rates.[58]

After the turbulent institutional attempts in the war and peace periods of the first half of the twentieth century, the second half saw the fading away of the Stalinist production system of the Soviet Union until the 1990s, and more recently the flourishing of a modified variant of this system in China. These basic breaks in global political economy set the frame for the institutional variants accommodating capital in the Western world since 1992. In particular, the evolution of financial institutions since that point in time can reveal interesting insights.

[56] See (Spencer 1862), a contemporary and rival of Darwin, who combines his evolutionism with the strict market liberalism [against any state intervention, (Spencer 1884)] so typical for aristocratic scientists of the nineteenth century.

[57] Compare (Radax et al. 2009).

[58] It is tempting to interpret the second phase, which is so dominant in Diagram 2, as being based on 'exploitation via flexible exchange rates in a globalizing world of TNCs'. The success of capital on a global scale in this second phase contrasted by only continuous growth in Diagram 1 (average GDP per capita) translates into an opening spread of incomes: Those derived from these global successes explode, while those (in Western Europe) which were not had to grow slower than before 1982 to arrive at the continuity shown in Diagram 1.

From an evolutionary perspective, such an investigation should take account of a too simplistic analogy, a mistake to be found in the concluding chapter of the economic historian Niall Ferguson's recent book (Ferguson 2008, pp. 341–358):

> "*Financial history is essentially the result of institutional mutation and natural selection. Random 'drift' (innovations/mutations that are not promoted by natural selection, but just happen) and 'flow' (innovations/mutations that are caused when, say, American practices are adopted by Chinese banks) play a part. There can also be 'co-evolution', when different financial species work and adapt together (like hedge funds and their prime brokers). But market selection is the main driver. Financial organisms are in competition with one another for finite resources. At certain times and in certain places, certain species may become dominant. But innovations by competitor species, or the emergence of altogether new species, prevent any permanent hierarchy or monoculture from emerging. Broadly speaking, the law of the survival of the fittest applies. Institutions with a 'selfish gene' that is good at self-replication and self-perpetuation will tend to proliferate and endure*." (Ferguson 2008, pp. 350–351)

Though it might be a kind of excuse that the text was written in May 2008, when the full extent of the looming crisis of finance institutions what not yet visible, it nevertheless is a timeless example of undue transplantation of biological metaphors. Financial institutions are not born by simple 'innovation/mutation'; they are strongly linked to the regulatory rule system of the nation states and thus are shaped in a way that reflects the surrounding political entity. The death of these finance institutions usually is not caused by competitive rivals in markets for 'scarce resources'. It rather needs a complex procedure—including several political entities sometimes, from state agencies to unions—to allow a large financial intermediary to go bankrupt. Finally, Richard Dawkins' suggestion[59] to re-introduce a moral concept, selfishness (remember Mandeville's 'vices'), does not lead to a better understanding of the performance of financial institutions—to say the least.

As argued in this paper, instead of unwise direct use of concepts of evolutionary biology, the evolution from money to capital is a process that evolutionary economics has to explain as an indispensable part of the economic evolution in general, of monetary political economy. The two important trajectories shaping the last 200 years of capital were technological progress and institutional evolution, and it is in the light of their development that financial institutions have to be understood. The evolution of the state, again, is a rather big topic; only a selected range of issues can be mentioned here.

The 'state' today comes mainly in three formats[60]: As nation state, as continental unit, and as a globally governing political unit. In the USA and in most parts of the EU, the elementary money form (US Dollar, Euro) is provided at the continental level, in Asia, such a unit is still in the making. The most developed form, the

---

[59] Compare (Dawkins 1989), and see (Lewontin et al. 1984) for a critique of Dawkins'attempt to extend biological metaphors to the social sciences.

[60] The focus on these three formats is due to the emphasis on the state's role as a monetary authority. A broader approach centering on the many ideological tasks the state takes care of can be found in (Althusser 1970).

capital algorithm, in principle works on the global level, though the level of regulations for most rules is the national level. Financial institutions carrying and executing the regulations are to be found on all three levels, and there is a trend to move power to the upper levels, leaving smaller tasks to subsidiary lower levels.

With respect to the existence of respective political bodies, their evolution on all three levels, and in particular towards the highest level, entered a hot phase since 1982. The emergence of the EU is a very recent development, a similar entity uniting China, Japan, and India has not even started—not to speak of the global level. In each of these institutional evolutions, specific history and culture of the concerned area—formally spoken: foregone path dependency—play a crucial role. As a consequence, ***evolutionary economics has to revive its history component*** to understand better current options and possible future trajectories of institutional evolution. In a sense, the current general crisis is just a symptom of the mismatch of the available institutional carrier systems and the capital program of large scale private carriers. Given this interpretation, the depth of the crisis is not surprising—hopefully it ***gives birth to a new global institutional design***. Many current proposals of singular adjustments of some rules of the prevailing financial architecture (e.g. Basel 3) miss this point and will prove as useless as will be the desperate conjurations of ethical behavior of capital managers. This crisis is ***not*** about the misbehavior of individual physical persons.

What has to fit a new global institutional setting is a worldwide ***structure of production units***, a structure that builds on (and partly transforms) the existing structure.[61] The separation of functions between globally acting TNCs and local SMEs, the main providers of employment, is already taking place.[62] The original role of the banking system in the twentieth century—namely, to discover the most profitable routes for further capital expansion and to channel credit towards these investments—has stagnated in the last 2 decades. Profitable intrusion of new spaces for investment became difficult. Indeed, the financial hype after the IT bubble in 2001 was built with the help of self-fulfilling prophecies, which needed ***not*** to be grounded on actually existing expectations concerning profitable real investment[63]—simply because such possibilities became extremely rare. It is thus a logical consequence that the mismatch between the most advanced money form (the capital program) and its welfare enhancing 'historical mission' first made its appearance as the fall of a carrier system of that mission, the fall of a large financial institution, the fall of Lehman Brothers on September 15, 2008.

Immediately after this pivotal turn, the role of state intervention—on all three levels and in content contrary to almost all policy discussion since 1982—was

[61] The contemporary interaction between TNCs and nation states concentrates on choice of capital structure of the former given 'political risks' set by the latter, as empirically studied by e.g. (Kesternich and Schnitzer 2009).

[62] In a recent contribution, this development of global firm structure and its connection to global mega-cities is discussed in more detail (Hanappi 2009).

[63] A more theoretical companion paper to this paper discusses four reasons for the financial crisis in more detail (Hanappi and Rengs 2008).

suddenly on the agenda again.[64] Since then, the debate more and more concentrated on the question of where and how the state should intervene, and if it should intervene at all was discussed less and less. But if the necessity for the state's active role is taken for granted, then the next questions are for the type of intervention and for quality and democratic control of the state' s decision makers (again at all three levels). But these questions are nothing other than the above mentioned call for the design of a desirable institutional setting.

5 Some Policy Conclusions

Any policy conclusion—in particular in critical situations—rests on a vision of possible future developments, in a less mundane language: on forecasts. Forecasts in times of deep crisis need a far reaching interpretation of the past (e.g. the one given in the previous sections of this paper) as well as some empirical evaluation of what was just happening in the immediate past. There exists an impressive flood of descriptions concerning the latter[65]; Diagram 3 just shows the recent development of the Dow Jones Index.

To make sure that this rather dramatic short-run development in the world's stock exchanges is not just the usual working of these markets necessary to clean it from unsound expectations, take a look at the long-run, real GDP growth rates of the USA, Germany, Austria and the four next largest countries of the Eurozone (France, Italy, Spain, UK) in Diagram 4.

As the diagram shows, there can be no doubt that this is the worst crisis for output and employment in the world since the Great Depression. According to our (continuously improved but always preliminary) forecasts, the worst is still to come: The fall in employment has hit all countries and there is no prospect for recovery.[66] How far average household incomes did fall due to this employment crunch depended heavily on the incomes policy of the respective nation state. By subsidizing wages and supporting private and semi-private financial intermediaries (e.g. banks and insurance companies), many countries translated the global crisis into a national public debt crisis. Given these measures, most European households were able to keep consumption at only slowly decreasing levels—using up their savings and credit worthiness. But then, starting with the case of Greece, international short-run speculation discovered the use of large scale profit rate expectations derived from producing information about possible public debt default of a country.

---

[64] As a recent contribution, compare (Auerback 2009).

[65] One of the most informative papers is (Calomiris 2008), a survey of events is (Furceri and Mourougane 2009).

[66] See (Wray 2009) for an appraisal of the importance of avoiding high mass unemployment. Martin Shubik, in a similar vein, suggests social innovation in the form of a 'Federal Employment Reserve Authority' Compare (Shubik 2009).

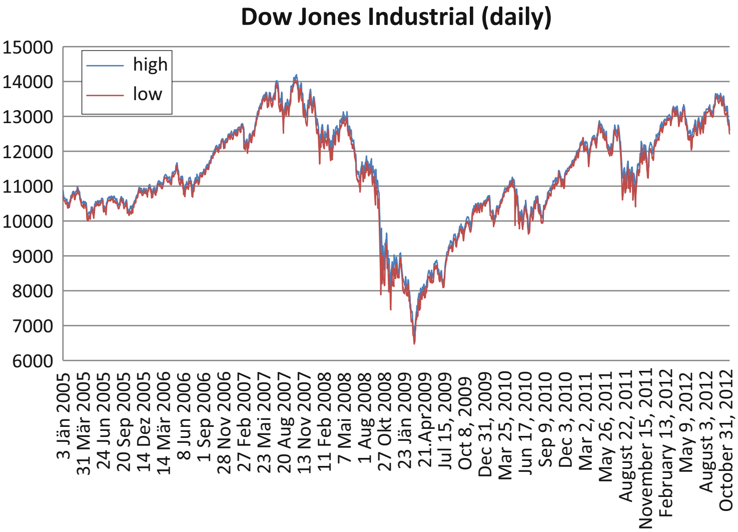


Diagram 3 Dow Jones Index (daily) as a description of recent events

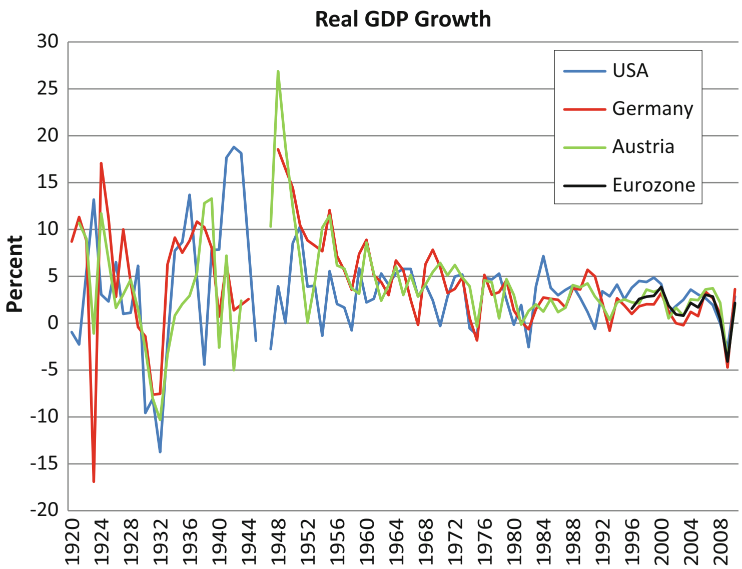


Diagram 4 Growth rates of real GDP

In the last years, this new game of international finance gained momentum, spreading to all Mediterranean countries and Ireland, and called into question the very existence of the European Union. Indeed the EU seems to be at the crossroads: Either it falls prey to the global capital program, which aims at restoring profitability in the weakest parts of the European country chain by re-introducing exploitation standards which European citizens today only know from TV-programs about less developed parts of the world. Or Europe unites and constitutes itself in a much more stringent and consistent way, as currently is the case. To avoid European disintegration, a close look at the reasons for disintegrative forces, an analysis of the emerging contradictions, will be the first necessary step. This immediately leads to the incompatibility of the globally working finance capital program and full employment in Europe.[67] To some extent, the notion of full employment might be redefined by reforming labor time organization. But still many questions remain: How can European public political institutions be implemented, which can secure the existing welfare standard at a zero growth rate? What can be the position of European economic activity in the global division of labor? If political institutions are not forced to grow—and to pay interest—the new monetary regime in Europe must still be made compatible to the global financial architecture. Add the even more burning question that the chain reactions of the global crisis are finding other weak chains in the highly interdependent global political economy with accelerating speed, the upheaval in the Arab countries is just the latest outstanding example.

In that situation, social and political pressure on policy makers already is enormous. This might enhance the efforts to develop and to implement a new institutional setting on a global level. But it also is an extremely dangerous transition period—comparable in some respect to the 1930s of the last century—since the history-dependent national responses to such a crisis easily can lead to locally dangerous phenomena, to governments based on direct coercive power instead of democracy.[68] The availability of highly efficient military means makes such a threat indeed more threatening than ever.

But even if it is possible to avoid the worst,[69] there will be no return to a smooth working of the old capital algorithm with high profit rates. Indeed the two big bubbles (ITC and finance) at the beginning of the century showed that the recent surge of 5 boom years had been built on thin air. Of course, the omnipresent capital program will not vanish, but it will lose its dominant role, and in most advanced OECD countries will have to be content with very low accumulation rates. No contemporary TNC can be imagined that can keep its organizational structure and culture with such a low profit rate. So the global setup of firms, of production units, will have to change profoundly. The lower turning point leading to this mild

[67] Some more sanguine politicians still express their pointless hope for an automatic reconciliation of this contradiction as their ‘hope for growth’.

[68] Therefore, a sophisticated regulatory response of institutions on the national level, tailored to the respective situation, is particularly important. This, of course, includes financial institutions; see (Nier 2009).

[69] Olivier Blanchard suggests a set of Keynes-style, demand-oriented measures to fight the crisis; Barry Eichengreen concentrates on regulatory remedies engineered by a reformed IMF; see (Blanchard et al. 2008) and (Eichengreen 2009). An alternative (even more ‘Keynesian’) interpretation and proposal for a policy recipe is provided by Jan Kregel (Kregel 2009).

recovery might be reached in 2015, so reorganization should start within the mid-term planning horizon of large firms. But there will be some firms winning, at least in terms of market shares of the shrinking market, and they will be hard to convince to accept new global political rules.

In several other aspects, a return to direct political measures can be expected, too. To secure a sustainable level of employment labor time regimes will be adjusted rather quickly—short time work will be here to stay and solutions tailored to the needs of production units and households hopefully will be found. Direct employment as state employees will start to play the important role of an immediate remedy. Since all this increases government expenditure, the question of who is lending to the government has become virulent. If the crisis is global, the traditional banking sector will not be able to step in and a partial return to measures reminiscent of command economies—perhaps including a moratorium on public debts in the USA and Europe—will be inevitable. As in wartime, a war against unemployment lead and 'financed' by governments might be a preferable solution.

With respect to money, this highly speculative outlook strongly points to a new metamorphosis of the money form. As was the case with previous form changes, the new form will ***not*** simply replace the current one; it rather will superimpose new and dominant features.[70] These features will have to be able to solve the most pressing problems of the global human species in the new century. The global character of these bottlenecks clearly calls for a large scale political decision procedure—democracy at work—carried out and executed by highly professional and committed agencies.

Global economic policy will have to work along the lines of the respective next bottleneck. The major three problem areas to be solved next—the new 'historical mission' of the new money forms following the capital algorithm—could be:

(1) Solving the question of avoiding a fallback to regimes of direct coercive power
(2) Solving the questions of sustainable environmental conditions
(3) Solving the questions of increasing income and wealth inequality in the world

Channeling economic and political activity towards work solving these problems will again be a program rather than a simple sign system of social values. It will even be a more complicated program than the capital algorithm residing in business plans and psychological traits of humans today. As a top-down program, it will need not only conscious planning but even planning of consciousness of the society as a whole. It thus will be a new experience for mankind. Money, the materialized expression of social value, on its long journey to ever greater abstraction, will start to become intentional on a global, social scale. Technical progress

[70] This, of course, is also true for technical innovation proper, for new products and processes. Technical innovation will go on, but the newly dominating form of reproductive innovation—out of need—will be more and more ***social innovation***. Technological innovation will turn towards ***change*** of products and processes, capital ***growth*** as a motive for innovation will fade out. The entrepreneur as social entity will be transformed along these lines.

has already produced the preconditions[71]; institutional evolution—indeed the implementation of adequate mechanisms of democracy—is knocking at the door of scientists, of evolutionary economists, to complete a synthesizing vision for this century.[72]

Bibliography

Althusser L (1970) Idéologie et appareils idéologiques d'Etat, La Pensée, no 151, juin 1970

Amable B (2003) The diversity of modern capitalism. Oxford University Press, Oxford

Auerback M (2009) The return of the state: the new investment paradigm, The Levy Economics Institute Working Paper No. 561

Blanchard O, et al (2008) Fiscal policy for the crisis, IMF Staff Position Note, December 2008

Brakman H et al (2007) Nations and firms in the global economy. Cambridge University Press, Cambridge, MA

Braudel F (1986) Sozialgeschichte des 15.-18. Jahrhunderts, Aufbruch zur Weltwirtschaft (Civilization matérielle, économie et capitalisme. XV[e] – XVIII[e] siècle, Paris, 1979). Kindler Verlag, München

Calomiris Ch (2008) The subprime turmoil: what's old, what's new, and what's next. Paper prepared for presentation at the Federal Reserve Bank of Kansas City's symposium, "Maintaining Stability in a Changing Financial System,"in Jackson Hole, Wyoming (USA), 21–22 August 2008

Catephores G (1994) The imperious Austrian: Schumpeter as bourgeois marxist, New Left Review I/205, May–June 1994, pp 3–30

Comte A [1979 (1844)] Rede über den Geist des Positivismus (französisch-deutsch). Felix Meixner Verlag, Hamburg

Davies G (2002) History of money. From ancient times to the present day. University of Wales Press, Cardiff

Dawkins R (1989) The selfish gene. Oxford University Press, Oxford

De Brunhoff S (1973) Marx on money. Urizan Books, New York, NY

Eichengreen B (2009) Out of the box thoughts about the international financial architecture, IMF Working Paper 116, May 2009

Elias N (1993) Time: an essay, 1st edn. Blackwell, London

Elsner W, Hanappi H (eds) (2008) Varieties of capitalism and new institutional deals. Edward Elgar, Cheltenham, UK

Ferguson N (2008) The ascent of money. A financial history of the world. Penguin Group, Australia

Foley D (2006) Adam's fallacy. Harvard University Press, Cambridge, MA

Frank G (1978) Dependent accumulation and underdevelopment. Monthly Review Press, New York, NY

Furceri D, Mourougane A (2009) Financial crises: past lessons and policy implications, OECD Economics Department Working Papers no. 668, Feb. 2009

[71] In a large research project for the Austrian National Bank, the technological impact on money forms has been studied (Hanappi 1999).

[72] It is encouraging that the Nobel prize in 2009 was given to Elinor Ostrom, whose major contributions—contrary to many journalists' perceptions—concern the 'understanding of institutional diversity' (Ostrom 2005).

Goodwin R (1955) A model of cyclical growth. In: Lundberg E (ed) Business cycle in the postwar world. Macmillan, London
Hanappi G (1989) Die Entwicklung des Kapitalismus. Peter Lang Verlag, Bern
Hanappi H (1994) Evolutionary economics. The evolutionary revolution in the social sciences. Avebury, Aldershot
Hanappi H (ed) (1999) Die Zukunft des Geldes, Wiener Studien zur politischen Ökonomie, Band 13, Wien
Hanappi H (2006) Endogenous needs, values and technology. Evolutionary economic modelling to replace microeconomics and macroeconomics, proceedings of the European Association for Evolutionary Political Economy (EAEPE) Conference 2006 in Istanbul (Turkey), Istanbul, 2–4 November 2006
Hanappi H (2009) Global cities and global firms? On the links between trends in urbanization structures and production structures. Global Bus Econ Anthol 1:41–48
Hanappi H, Rengs B (2008) On the eve of global financial collapse. Explanations, necessity, and policy conclusions, proceedings of the International Conference of EAEPE in Rome (Italy), 6–8 November 2008
Harrod R (1939) An essay in dynamic theory. Econ J 49:14–33
Hayek F (1931) Prices and production. A. M. Kelley, New York, NY
Hegel GWF (1807) Phänomenologie des Geistes, Jena
Heinsohn G (1984) Privateigentum, Patriarchat, Geldwirtschaft. Suhrkamp, Frankfurt am Main
Hilferding R (1910) Das Finanzkapital. Europäische Verlagsanstalt, Wien
Hobsbawm E (1996) The age of extremes: a history of the world, 1914–1991. Vintage Books, London
Hodgson G et al (2001) Capitalism in evolution. Edward Elgar, Cheltenham,UK
Ilyina A, Samaniego R (2009) A multi-industry model of growth with financing constraints, IMF Working Paper 119, May 2009
Jevons S (1871) The theory of political economy. Macmillan, London
Kesternich I, Schnitzer M (2009) Who is afraid of political risk? Multinational firms and their choice of capital structure. Discussion Papers of the Deutsche Bundesbank no. 02/2009
Keynes J (1936) The general theory of employment, interest and money. Macmillan, London
Kondratiev N (1926) The long waves in economic life, Archiv für Sozialwissenschaft und Sozialpolitik (transl. 1935, REStat)
Kregel J (2009) Managing the impact of volatility in international capital markets in an uncertain world, The Levy Economics Institute, working paper no. 558, April 2009
Krugman P (1996) The self-organizing economy. Blackwell, Cambridge, MA
Lechevalier S (2007) The diversity of capitalism and heterogeneity of firms—a case study of Japan during the lost decade. Evol Inst Econ Rev 4(1):113–142
Lewontin R et al (1984) Not in our genes. Biology, ideology, and human nature. Pantheon Books, New York, NY
Machiavelli N [1988 (1532)] The prince, Cambridge texts in the history of political thought, Cambridge University Press, Cambridge, MA
Maddison A (2006) Contours of the world economy, 1–2030 AD. Oxford University Press, Oxford
Mandeville B (1714) The Fable of the Bees or Private Vices, Publick Benefits, 2 vols. With a Commentary Critical, Historical, and Explanatory by F.B. Kaye (Indianapolis: Liberty Fund, 1988). Vol. 1
Mankiw G (2010) Macroeconomics, 7th edn. Worth, New York, NY
Marx K (1848) Das kommunistische Manifest. In: Marx-Engels Werk, Dietz Verlag, Berlin, 1974
Marx K (1857) Das Kapital (Band 1). In: Marx-Engels Werke Bd. 23, Dietz Verlag, Berlin, 1974
Mendoza E, Quadrini V, Rio-Rull J (2009) Financial integration, financial development, and global imbalances. J Polit Econ 117(3):371–416
Menger C (1871) Principles of economics. Wilhelm Braumiiller, Vienna
Mishkin F, Eakins S (2011) Financial markets and institutions, 7th edn. Prentice Hall, San Francisco, CA

Baron de Montesquieu (1748) The spirit of the laws (1912). D. Appleton and Company, New York. http://web.archive.org/web/20080821125602/http://etext.lib.virginia.edu/toc/modeng/public/MonLaws.html, De l'esprit des lois, translated by Thomas Nugent
Müller RW (1977) Geld und Geist. Campus-Verlag, Frankfurt
Nier E (2009) Financial stability frameworks and the role of central banks: lessons from the crisis, IMF Working Paper 70, April 2009
Ostrom E (2005) Understanding institutional diversity. Princeton University Press, Princeton, NJ
Quesnay F (1758) Le Tableau Économique, Paris
Radax W, Wäckerle M, Hanappi H (2009) From agents to large actors and back. Formalized story-telling of emergence and exit in political economy, paper contributed to the 11th conference of the Association for Heterodox Economics, London, June 2009; forthcoming in the proceedings
Schumpeter J (1939) Business cycles: a theoretical, historical and statistical analysis of the capitalist process. McGraw Hill, New York, NY
Shubik M (2009) A proposal for a federal employment reserve authority, The Levy Economics Institute of Bard College, Policy Note 2009/5
Smith E, Foley D (2002) Classical thermodynamics and economic general equilibrium theory. Santa Fe Institute, Santa Fe, NM
Sohn-Rethel A (1978) Warenform un Denkform, edition suhrkamp, Frankfurt a. M
Spencer H (1862) First principles. Williams & Norgate, London
Spencer H (1884) The man versus the state. Williams & Norgate, London
Thomson G (1972) The first philosophers: studies in ancient Greek society. Lawrence & Wishart, London
Veblen T (1899) The theory of the leisure class. University of Chicago, Chicago, IL
Walras L (1874) Elements of pure economics, or the theory of social wealth. Lausanne, Paris
Walsh C (2010) Monetary theory and policy, 3rd edn. MIT, Cambridge, MA
Wray L (2009) The social and economic importance of full employment, The Levy Economics Institute, working paper no. 560, April 2009
Wright EO (ed) (2005) Approaches to class analysis. Cambridge University Press, Cambridge, MA

Asymmetric Markets and the Evolution of the Division of Labor

Carl Christian von Weizsäcker

Abstract In a framework of generalized Darwinism (selection, variation, replication), freedom of choice by the individual is an important ingredient with respect to all three criteria. Such an evolutionary economic system rests on market asymmetry, where the supply side consists of firms that compete and the demand side consists of individuals or firms that choose without being under competitive pressure.

1 Introduction

Do we believe in the introductory sentence to economic science? It reads, as you know: "The greatest improvement in the productive powers of labour, and the greater part of the skill, dexterity, and judgement with which it is anywhere directed, or applied, seem to have been the effects of the division of labour." (Adam Smith, Wealth of Nations, Book 1, Chap. 1, First Sentence.) If we do, and if we work as evolutionary minded economists, we might be interested in the evolution of the division of labour. In this paper, I want to point to a particular feature of that evolution, which to my knowledge has not been treated before in the economic literature. I have a specific reason to present this idea in the Darwin year 2009. The paper is an attempt to understand the concept of competition, which is, of course, embedded in Darwin's concept of "natural selection".

C.C. von Weizsäcker (✉)
Max Planck Institute for Research on Collective Goods, Bonn, Germany
e-mail: weizsaecker@coll.mpg.de; ccvwzs@t-online.de

G. Buenstorf et al. (eds.), *The Two Sides of Innovation*, Economic Complexity and Evolution, DOI 10.1007/978-3-319-01496-8_14,

2 Competition and Choice: Selection in the Production Sector

In economics, we have the following idea: consumers, by their demand on markets, give orders to producers as to which goods to produce. This is the idea of consumer sovereignty. Competition among producers is supposed to serve this principle of consumer sovereignty. The consumer is the boss and the producers, under the pressure of competition, are his underlings. In deciding what to consume, the consumer follows his preferences. He is free in his choice, subject to a budget constraint. (The concept of "preferences," in my opinion, is the way economics introduces the idea of freedom of choice into the formal model of the interaction of individuals. I do not discuss this interpretation of preferences in the present paper). In particular, the consumer is free as to where to buy, and which of the competing suppliers to patronize. This freedom of choice or freedom of selection imposes competitive pressure on suppliers, whereby the production process at large follows the instructions coming from the consumers. In this way, the preferences of the people are mirrored in the production programs of the producing part of the economy.

There is, then, a market asymmetry between supply and demand. One way practitioners of competition policy use this asymmetry is by using the concept of the relevant market in finding out who competes with whom. Product markets in the modern economy are characterized by a particular feature: at the going competitive price, suppliers are demand constrained, whereas customers can buy as much as corresponds to their needs, given their income and given prices in this market and in other markets. Another way to express this is to say that, after the transaction in the market, the customer is "transaction saturated", whereas the supplier remains "transaction hungry". The supplier would have liked to have sold more than he actually succeeded in selling, as a rule. This is also related to the fact that, in routine purchases (small item purchases), the seller sets the price and the customer determines the quantity. The price set by the supplier is above his marginal cost even in a state of normal competition. Given quality, the gap between price and marginal cost equals the cost of the marginal selling effort.

Commercial competition, then, means the following: If the customer decides to place an order with supplier A, then supplier B does not get that order. If supplier B had supplied that customer before then, supplier B has "lost" this customer to A. Other things being equal, it means that supplier B now sells less than he would have sold had the customer stayed with him. Other things (prices, selling effort, quality etc.) remaining equal, supplier B cannot simply "replace" the lost customer c by another customer d who previously had not been his customer. This is what we mean when we say that A and B compete.

On the demand side, this is different. Customer c can "replace" supplier B by taking orders to supplier A. This presents no problem (except in the presence of switching costs, but for the present simplified description I ignore these). So, it does not make very much sense to say that customer c and customer d "compete". In a normal state of the market, supplier A happily serves both customers; after all, he is transaction hungry. The fact that my neighbor has bought an Audi car from his dealer does not prevent me from also buying an Audi from the same dealer. But the fact that I bought an Audi from the Audi dealer now means that the BMW dealer cannot hope to sell me a BMW.

If the relevant market, in which, say, supplier A operates, is to collect those suppliers who compete with A, then we implicitly assume that many customers of A—not necessarily all of them—have the choice to replace A by one of those other suppliers in the relevant market. The other side of the relevant market is the side that has the choice. These customers then do not compete. If customers would have to compete, and if there is also competition on the supply side, the concept of a relevant market in its present anti-trust meaning would make no sense. Indeed, assume a "market" in which suppliers compete for customers and, at the same time, buyers compete for suppliers. Thus, if supplier A had so far supplied customer a, supplier B had so far supplied customer b, supplier C had so far supplied customer c, then a switch of supplier by customer a from A to B, given the fact that customers also compete, can only be accomplished, if customer b is pushed out by supplier B. In other words, the "choices" by customers only can be made interdependently. Then the thought experiment—what happens to demand if supplier A raises his price by 10 %—does not make sense, or, rather, is unbelievably complicated. The usefulness of defining a relevant market, say by means of a SSNIP-test (small but significant and nontransitory increase in price), disappears.

The concept of the relevant market, then, is tied to the beautiful simplicity of transactions in those markets, generated by the fact that any individual on one market side can choose independently from the choice of others on the same market side. This presupposes that the other market side, the market side in which people compete, is transaction hungry. A simple formula: one market side chooses, the other market side competes.

We may relate this to the "and/or—distinction". If the supply side competes and the demand side chooses, customers take supplier A or supplier B or supplier C or . . .. Suppliers take customer a and customer b and customer c and . . ..

If we work in a scheme of "generalized Darwinism" (Aldrich et al. 2008) the traditional concept of competition among producing firms corresponds to the "selection" part of that scheme. We may also understand firms as "programs" in the sense of Ernst Mayr, which provide the action and reaction habits of individuals in their adaptation to their respective environment. The choice action of the consumer, then, does not belong to the competitive part of the scheme. There is no selection among consumers. The consumers are part of the environment within which the firms adapt. The consumers are part of the environment that "selects", providing the criteria according to which the selection process in the production sector operates.

This description of market asymmetry would be incomplete without an understanding of the effects of the selling effort by suppliers. Whereas competition restricts the options of suppliers, consumers are restricted in their choice by their limited information about the availability of products and their prices. This allows suppliers to direct the choice of consumers beyond the pricing of their products, by informing or otherwise influencing customers. Economists have debated the usefulness of advertising and other forms of selling effort. Whatever the conclusions concerning the social benefit of suppliers' selling effort, it is clear that such effort does have an influence on consumer choice. Some critics have described this selling effort as a "manipulation" of consumers. Nevertheless, it remains true that, in typical markets, consumers have a choice and there is no competition among consumers.

Market asymmetry, as it works in practice, does not mean that the actual situation is an optimum. No doubt, improvements in the detailed institutional set-up are always possible. But, whatever these improvements, say, in the form of regulation or government sponsored consumer information, they will not fundamentally change the asymmetry characteristic of typical product markets. This characteristic is an outflow of the division of labor, without which living standards would be substantially lower than they actually are.

The other part of that environment is, of course, the availability of production technologies including the know-how embedded in the people who work in the production process, and the preferences of those people.

3 Innovation and Choice: Variation in the Production Sector

In the "generalized Darwinism" scheme the two other great principles are "Variation" and "Replication". Evolutionary economics has consistently pointed to innovation and, in particular, Schumpeterian innovation as the source of evolutionary variation. Attempts to innovate generate variation; successful innovation then provides the evolutionary direction, which is proven by the process of selection that economists call "competition". In my analysis, I follow this line of thinking, although I want to add that consumer choice is an important ingredient of this evolutionary process.

This can easily be seen by making the thought experiment to the contrary: assume for a moment that the demand side of markets, in particular consumer demand, would also be characterized by competition. As I have shown in the preceding section, this would mean that any individual demand no longer would be independent from the demand of other individuals. But this would also mean that changes in demand would become unbelievably complicated. The system would be condemned to stagnation, to an adherence to the status quo. Thus, innovation attempts strongly depend on the individual independence of consumer demand, or more generally, independence of demand. Since competition implies interdependence, it means that consumer choice is free from competition, and so freedom from the pressure of selection is an important ingredient for variation within the system, i.e. an important ingredient for innovation and thus evolution.

The principle of asymmetric markets, of competition on the supply side and choice on the demand side, then, is a necessary condition for an innovative economy.

We may say the same thing in different words: how is it that our human society has been able to generate an economic world of such complexity? The answer is: by relying on a mode of transaction of extreme simplicity. The merchant sets the price (in a competitive environment) and the consumer buys the quantity he wishes. This allows buyer diversity: due to low transaction costs, the consumer can be on the demand side of so many different markets. Thus the production sector is able (and under competitive pressure is forced) to differentiate itself into many different markets, each of which is characterized by highly specialized suppliers. Thus the simplicity of market transactions due to the market asymmetry is the precondition for the high degree of complexity of the system as a whole which thereby can reap the benefits of a high level of division of labor. Moreover, a change in the system can be accomplished with relative ease due to the same simplicity of transactions: a new product, a new market can be launched, because consumer demand is individually independent, i.e. is not under the pressure of competitive interdependence.

The same of course is true for the other prerequisite of innovation in the production sector: relative ease of access to the factor inputs required for innovation. Resources are available which the innovator can buy on the open market. This aspect of innovation has been studied extensively under the heading of "Silicon Valley" and innovation networks.

4 Progress, Efficiency and Choice: Replication

"Natura non facit saltum" is the motto which Alfred Marshall, one of the pioneers of evolutionary economics, put in front of his "Principles of Economics". As we look at the process of economic evolution, it is very fast in comparison to most evolutionary development in the world of plants and animals, but it is rather slow, if you simply conduct your daily life within the economic system. Indeed, our ability to pursue our individual goals heavily depends on the fact that things today are not very different from things yesterday, and we can expect things tomorrow to be very similar to things today. If this were different, we would have great difficulty finding our way through the economic world. Thus, we obviously can rely on a good functioning of the third principle of "generalized Darwinism": replication. Here I do not discuss the detailed mechanisms of replication, which are, of course, related to biological replication: younger people take up the jobs of older people who retire; and schooling experience can rely on similar natural abilities of the young as compared to the old. (Schooling is, of course, an important part of the replication mechanism in modern society.) Then there are "social habits" in families, nations, firms, schools, etc., which are transferred by instinctive more than rational imitation. "Culture" in a broad sense of the word encompasses many mechanisms of replication.

What I want to add to the "received wisdom" of evolutionary economics is the following observation. I believe there is a misunderstanding between the evolutionary minded school and the orthodox (neoclassical) school in economics. This misunderstanding concerns the concept of "efficiency". There are two quite different meanings of this term: Pareto-optimality and Kaldor-Hicks-Scitovsky efficiency. The concept of Pareto- optimality is rather useless for the analysis of evolving economies. In this paper, I do not show this in detail. By contrast, the concept of Kaldor- Hicks efficiency can be very useful, although it strongly depends on Marshall's "natura non facit saltum". Pareto-optimality is a concept that does not involve market prices, while Kaldor-Hicks efficiency depends on the use of actual market prices: it asks whether a change provides more in terms of real income gains than in terms of income losses, and to measure the change in real income, it uses actual market prices. The concept is useful, if it is applied to changes from the status quo, which are "small" relative to the economy at large, so that most relative prices are not affected by the change. Therefore, its usefulness relies on the "natura non facit saltum" principle, which, by and large, can be assumed to hold.

Assuming "natura non facit saltum", we can ask questions: under which conditions is it the case that changes in the economic system are in all likelihood efficient? Do we have a filter that selects changes in accordance with the principle of efficiency? These kinds of questions relate the evolutionary process to the normative idea of improvement or progress.

One of the important ingredients of an "efficiency filter" for changes in the economy is working property rights, which makes it more difficult to obtain changes without the consent of the people affected by the change. Property rights, for better or worse, are partial veto rights against changes in the economic set-up. This leads to the idea of an optimal density of property rights. Too little density implies that many inefficient changes can take place, changes against the will of many people who are more or less affected by the change. Too high a density of property rights makes change altogether difficult and thus does not only veto *inefficient* changes but also *efficient* changes.

The efficiency criterion rests on the evaluation by individuals according to their preferences. Therefore, it rests on the idea of free choice by individuals. Choice by individuals then is involved in all three characteristics of the economic evolution in a market economy. It is the appropriate counterpart of the competitive process in asymmetric markets, which guides the production sector in accordance with individual preferences (selection). It is the ingredient of the innovative process to enable variety and thus, in the long run, complexity by means of the simplicity of transactions in asymmetric markets (variety). It is the criterion for efficiency and thus the progress of gradual change in an economy which looks today almost as it looked yesterday (replication).

5 Mutual Adaptation of the Division of Labour and of Individual Choice

Innovation and progress by efficient change are well known topics in economic theory, be it evolutionary, institutional or more traditional in the neoclassical sense. Nevertheless there is very little literature that explicitly discusses the role of the division of labor in this process. The production process uses and reproduces patterns of know-how. There is then a mutual adaptation of the production process and the process which generates the distribution of know-how across the working population. But there is "drift" in the process that generates know-how patterns. It seems that the division of labour is increasing through time. There are plenty more innovations that allow useful human knowledge to increase than there are innovations that allow useful human knowledge to decline. Greater and greater specialization seems to be (on average) the course of economic history.

In parallel, the number of markets for goods made by specialized suppliers seems to increase through time. Does this process of growth of the number of markets come to an end at one time? Can consumers cope with this ever increasing diversity of markets? We do not know. But we should be aware that, as a basic principle, vertical specialization pushes up the limits of diversity with which end-consumers can cope. Integrators or aggregators can put together packages of different and diverse inputs into new goods, so that the end-user faces directly only a fraction of the goods involved in the total production process leading up to the goods consumed. These details of the evolution of the division of labour have not been studied so far.

Traditional welfare economics has dealt with normative questions mainly in the context of neo-classical theory. If we want to look for the normative issues of the evolution of the division of labour, it appears to be unlikely that the traditional instruments of welfare economics are sufficient. In particular, welfare economics has put aside the problem of a feedback from the production and consumption process to preferences. Yet it is difficult to imagine that preferences for goods remain unaffected by the goods that become available in the economic evolution. Much more plausible is the hypothesis that preferences also adapt to the economic process of evolution. Some people talk about habit formation. We then would see the economic process of evolution as a process of mutual adaptation of the technology and the division of labour to any given set of preferences at any one time and of preferences to any given state of technology and any given state of the division of labor.

It is a great intellectual challenge to find out whether concepts such as progress or improvement still can be maintained, if the measuring rod of economic performance (individual preferences) itself is influenced by or adapted to the economic process that is supposed to be gauged by it. Can we then still maintain our criterion of individual choice, which was so useful in understanding the evolutionary process?

6 Conclusion

We then see that freedom of choice of consumers and, thus, absence of competition among consumers in any particular market is a necessary structural ingredient of an economy that is built on the principle of the division of labour. The implied asymmetry of market structure is required to perform the “Selection” as well as the “Variation” part of the economy that conforms to the pattern of “Generalized Darwinism”. Moreover, “consumer sovereignty” also enables us to relay the “evolution” of the system to the normative concept of “progress”. And the progress concept is built on the idea of “Replication” that manifests itself in the Marshall’s motto: “natura non facit saltum”.

Reference

Aldrich HE, Hodgson GM, Hull DL, Knudsen T, Mokyr J, Vanberg VJ (2008) In defence of generalized Darwinism. J Evol Econ 18:577–596

Recent Developments in Evolutionary Biology and Their Relevance for Evolutionary Economics

Karin Knottenbauer

Abstract The paper gives attention to the question of whether the development of evolutionary theories in biology over the last 20 years has any implications for evolutionary economics. Though criticisms of Darwin and the modern synthesis have always existed, most of them have not been widely accepted or have been absorbed by the mainstream. Recent findings in evolutionary biology have started to question again the main principles of the modern synthesis. These findings suggest that phenomena of co-operation, communication, and self-organization have been underestimated, and that selection is not the predominant factor of evolution, but only one among many. Thus, in evolutionary economics, the question is whether the popular variation-retention-selection principle is still up to date. The implications for evolutionary economics with respect to analogies, generalized Darwinism, and the continuity hypothesis are also addressed.

1 Introduction

In the last 20–30 years, there have been discoveries in evolutionary biology that, on the one hand, have led to the inclusion of new fields of research, such as embryology and ecology, and on the other hand, to the emergence of new factors of evolution and thus to a shift in the significance of chance, genes and selection. The relationship of these findings towards the modern synthesis—the standard theoretical paradigm of evolutionary biology—is not yet clarified and agreed upon in the scientific community, but the new empirical evidence is so overwhelming that a new assessment of the theory of evolution is inevitable. The reason why this development could be interesting for evolutionary economics is that some approaches in evolutionary economics draw upon evolutionary biology, mainly

K. Knottenbauer (✉)
RWTH Aachen, Aachen, Germany
e-mail: Karin.Knottenbauer@rwth-aachen.de

G. Buenstorf et al. (eds.), *The Two Sides of Innovation*, Economic Complexity and Evolution, DOI 10.1007/978-3-319-01496-8_15,

through the form of metaphors and analogies or in unifying concepts such as generalized Darwinism. Thus, since evolutionary economics itself is a research program outside the mainstream, and since there are obvious similarities between the modern synthesis and neoclassical economics (Foster 2001, p. 118), it would be inattentive not to consider—or at least not to have a look at—these recent developments outside mainstream evolutionary biology.

Charles Darwin's intention was to provide evidence of evolution being a historical matter of fact and to explain evolution. Darwin's own explanation of the rise, change and decline of species was complex: in some respect clear and distinct, but in others vague, even contradictory, and since genetics had not yet been established, some explanations remained speculative. At the beginning of the twentieth century, with the rise of genetics, the explanation of evolution was still not generally accepted, as may be seen, e.g., in the scientific controversy between the geneticists and the researchers on populations about the relative significance of discrete mutations, the impact of the environment and selection. It was only later in the 1930s and 1940s when the main principles of genetics had become more widely accepted that the pluralistic view narrowed down to that of the modern synthesis (or synthetic view of evolution), which is with modifications the standard theoretical paradigm in evolutionary biology today.

The basic tenets of the modern synthesis might vary slightly from interpretation to interpretation (e.g. Reif et al. 2000), but there is a kind of common quintessence on which most of the exponents would agree. According to the modern synthesis, stochastic changes in gene frequency (caused by mutation, recombination and gene flow) produce small, heritable variations in the phenotype. Some of the individual organisms are, then, better adapted to the environment, have a higher fitness and spread in the population through positive selection. The ongoing modification of the genome furthers the morphological evolution and the only factor directing morphological evolution is the selection of better adapted phenotypes (Wieser 1994, p. 160).

In the following, and to avoid any misunderstandings, I would like to emphasize that the modern synthesis is considered here as a theory on how evolution took place. The issue, then, is the explanation of evolution and not the historical record of evolution which from a scientific perspective in this paper is taken to be a matter of fact. Including the research areas presented in the following, there is no disagreement about conceptions of evolving species in general, and about the principle of common descent.

Challenges to the modern synthesis in the second half of the twentieth century have been manifold, e.g. the neutral theory of molecular evolution by Kimura (1955, 1983), today better known under the label of "gene drift", endosymbiotic theory (Margulis 1970), theories of internal selection and development constraints (e.g. Gould 1980; Maynard Smith et al. 1985), system theories of evolution (e.g. Riedl 1975), theories of discontinuous evolution, the punctuated equilibrium theory (Gould and Eldredge 1972, 1977), theories of self-organization (Kauffman 1993), theories of skin, group and species selection and the corresponding hierarchical theory of evolution (e.g. Maynard Smith 1964; Vrba and Eldredge 1984;

Eldredge 1985), the theory of autoevolution (Lima-de-Faria 1988) etc. Most of these critical approaches did not reject the modern synthesis completely, but acknowledged at least some of its achievements, while extending, qualifying or putting into perspective certain tenets of the modern synthesis and also adding new factors of evolution. A lot of the critique of the modern synthesis implied the predominant roles of (random) mutations, genes and selection in biological evolution. Some discoveries, such as gene drift, were absorbed by the modern synthesis, whereas others, such as the endosymbiotic theory and punctuated equilibrium theory, entered standard text books on evolution but still were presented mainly as side effects that completed the standard explanation without questioning the main tenets of the modern synthesis. The modern synthesis would definitely have remained the standard theoretical paradigm in evolutionary biology had criticism in the last decades ebbed away. But the contrary has taken place. More and more empirical findings and new hypotheses critical to the standard theories have come up in the last 25 years. They draw on new insights into molecular underpinnings, genetics, ecology and developmental biology, giving new support to the old criticism of the modern synthesis. Apart from the publications in scientific journals, a set of books have been released that summarize and assess these discoveries, among them are Carroll (2006), Gilbert and Epel (2009), Jablonka and Lamb (2005), Kirschner and Gerhart (2005), and Pigliucci and Müller (2010a).

Thus, in Sect. 2, I will present essential research programs and research findings in evolutionary biology. The choice has been made according to the principles of (1) novelty—which will roughly be defined as what has emerged in the last 25 years (a period that admittedly is to a certain extend arbitrary but the history of the theory of biological evolution is not the main focus of the paper), (2) divergence from mainstream, and (3) ongoing research activities in this field. It should be mentioned that many of the recent research findings to be presented here are not completely new, but have been either better confirmed or have become clearer, so there might be an overlap with earlier approaches. There is no claim that the list of recent approaches in evolutionary biology is complete, since the aim of the paper is not primarily to give a comprehensive record of the developments in biology, but to hint at some methodological consequences for evolutionary economics from a choice of unorthodox biological theories. In Sect. 3, an assessment of the findings mainly by biologists is presented. In Sect. 4, possible implications for evolutionary economics are given. Sect. 5 concludes.[1]

[1] Since I am an economist and not a biologist and have had no training in biology, I apologize for any remaining obscurities and errors, in particular in the sections dealing with biology. Any comments and corrections are very welcome.

2 Recent Developments in Evolutionary Biology: Findings

A major motivation for biologists to explore evolution further was the fact that some phenomena of evolution have still not been explained satisfactorily. Among them is the so called Darwin's dilemma, since this problem had already been raised by Darwin himself. The question is: How can complex organs or parts of the body like an eye, a limb or a wing be the result of a gradual selection process? An eye, for example, is a complex system of a lens, a retina, tissue, nerves and so on. If an eye is indeed the product of gradual selection, it has not developed as a whole, and all of a sudden. There must have been intermediate forms with which the individual organism was not only viable but which were also selected, since according to the modern synthesis, every intermediate form, including incomplete ones, must have shown a selective advantage. It is obvious, then, that this explanation of the emergence of complex new structures was not satisfying (Kirschner and Gerhart 2007, p. 16). It was shown that this deficiency of the modern synthesis lies in its disregard of the rules of transformation from genotype to phenotype. The modern synthesis concentrates on two research areas, genetic and phenotypic evolution, implicitly assuming that the space between genetic evolution and phenotypic evolution is not relevant (Müller 1994, p. 185) and not knowing whether the genetic changes responsible for big changes in form were the same as those for small variations within species (Carroll 2006, p. 284). Thus, the modern synthesis was blamed for only explaining what body forms can be maintained in organismal evolution, but not what forms are generated. It was reasoned that mutations only produce small variations from what already exists, but do not create new forms (Kirschner and Gerhart 2007, p. 23) and mutations are often the result of negative impacts such as radiation and hence destroy more than they create. In addition, selection is a mechanism that can only work on what already exists (Müller and Newman 2003, p. 3). Thus, a theory of how completely new organs, structures and body plans come into existence was still lacking.

This scientific challenge has given rise to a research program called evolutionary developmental biology (or informally "evo-devo"), which combines two research areas that had existed quite independently until roughly 1990: developmental biology and evolutionary biology (Müller 2005, p. 87). Developmental biology deals with the development of organisms and ontogenetic rules of the development of forms and structure, in particular, the origin and evolution of embryonic development. Evo-devo seeks to extend this research by concentrating on interdependencies between ontogenesis and evolution, in particular, the analysis of causal relationships between embryonic and evolutionary processes (Müller 1994, pp. 155–160).[2] One of the main questions posed by evo-devo is: Does

[2] According to Müller (2005, pp. 98–102), evo-devo comprises four research programs: (1) The Comparative Morphology Program (2) The Epigenetic and Experimental Program (3) The Evolutionary Developmental Genetics Program (4) The Theoretical Biology Program.

developmental (embryonic) evolution bias the produced phenotypic variations or constrain certain paths of evolution? (Futuyma 2007, p. 474).

Some exponents of evo-devo have concentrated on physical processes that guide the way in which cells organize organs and tissues. The stickiness, elasticity, and chemical reactions within and between cells affect the body plan of an organism (Pennisi 2008, p. 196). Since body plans have internal inertia, evolution is not completely arbitrary, but works around these stable body plans. Consequently, there is a multitude of development constraints. The modern synthesis implicitly assumes that genetic and morphological variation and evolution are highly positively correlated. However, results in molecular genetics show that this is not the case and, even from the complete knowledge of the genome, one cannot infer the anatomy of an organism. The ontogenetic formation of morphological structures is not determined by genes only, but is due to an interdependent activity of genes, cells, tissue and external factors (Müller 1994, p. 162).

One of the discoveries of evo-devo in contrast to the modern synthesis has been that the same or similar organs and structures in different animals (such as eyes, limbs, hearts, legs, wings etc.) have not been "invented repeatedly from scratch", which means they would have emerged from distinct changes in the number and sequence of genes (Carroll 2006, p. 132). Instead, similar organs and structures have "evolved by modification of some ancient regulatory networks under the command of the same master gene or genes" (Carroll 2006, p. 286). In other words, common and very old genes or master genes have existed latently in the genome of all animals for a very long time (since early evolution), and they are the reason for similar body plans and organs in very different animals (in different branches of the genealogical tree) (Carroll 2006, p. 71). Genes that code for body plans are called Hox-genes. These old and complex genes are activated in embryonic development. Hox-genes together with transcription factors (proteins that bind to DNA and turn transcription on or off) and signaling pathways (communicating cells leading to traveling proteins that induce changes in the cell and in gene expression) are considered to be elements of a genetic tool box (Carroll 2006, p. 74). Similar sets of Hox-genes and sometimes even identical single genes can produce a variety of body plans because the genetic tool box has been differently combined by the function of genetic switches (promoters, enhancers), regions of DNA that are typically located near the genes upstream on the same strand of DNA, which can turn genes on and off in reaction to transcription factors (specialized signal proteins). Genetic switches allow the same tool kit genes to be used differently and thus are a central factor in the creation of variety. As a result, these genetic switches are considered to be the "key actors" in embryonic development and evolution (Carroll 2006, p. 111). Though the discovery of Hox-genes was important, it was discovered that it is not a single Hox-gene that is responsible for one new organ or structure, but groups of gene switches and proteins that build networks that regulate the formation of whole organs and structures (Carroll 2006, p. 129). Changes in genetic switches trigger a shifting in the zones of Hox-genes and thus are responsible for all kinds of changes in the body form, including the creation of new species. Therefore, not the invention of new

Hox-genes caused evolution but the genetic switches triggered by ecological conditions worked on the genetic tool kit. According to this view, natural selection is not responsible for the creation of new body forms, but determines only which forms are actually been realized (Carroll 2006, p. 287).

Kirschner and Gerhart argue on similar lines with their hypothesis of facilitated variation. They say organisms restrain some components of their phenotype from certain changes while, on the other hand, other components are released for change. Restricted elements are called conserved core components and they consist of a sequence of several protein components. These core components have remained relatively intact through time, and have not been subject to gradual change but rather to discontinuous waves of change (Kirschner and Gerhart 2007, p. 299). Specific features of these core processes make them robust and flexible, e.g. weak regulatory linkage, modularity, and compartmentalization. These features have not changed in the course of time but have provoked regulatory changes (Kirschner and Gerhart 2007, p. 354). The greatest amount of change (since the Cambrian) is not due to changes within core components, but is due to regulatory changes of core processes. Among these regulatory changes are changes in the date, place, amount, and circumstances of gene expression, cell signaling, the role of Hox genes in embryonic development, the program for developing extremities, etc. As a result of these regulatory changes, core components are used in new combinations and in different amounts at other times and in other places. Under the new conditions, the flexible core components might show different performance and new phenotypes (Kirschner and Gerhart 2007, p. 300). Carroll argues along the same line: "At an anatomic level, multifunctionality and redundancy are keys to understanding the evolutionary transitions in structures" (Carroll 2006, p. 288) and Müller states "[t]he new tenets by evo-devo may be called "emergence" and "inherency" (Müller 2005, p. 106). Core components are adaptive, and have been selected due to their robustness in embryonic development and anatomic stability. Consequently, phenotypes that have emerged from regulatory modifications of core components are presumably less lethal, more viable and better adapted than completely new phenotypes. In this way, variation is facilitated and accelerated (Kirschner and Gerhart 2007, p. 301). Organisms combine a limited number of components to transform a small amount of random mutations into new phenotypic forms. However, organisms and not mutations are the protagonists in the realization of concrete body forms (Kirschner and Gerhart 2007, p. 354). This view also stresses the existence of established and approved components for innovation (Kirschner and Gerhart 2007, p. 303).

It is well known that genomes consist of DNA with coding sequences (genes in the narrow sense) and DNA with non-coding sequences, also called "dark matter of the genome" (Carroll 2006, Chap. 5). In the last decades, it has emerged that, in the non-coding sequences—the larger part of the genome—specific elements, called "transposable elements" (or transposons) can change the structure of the genome, including changes to germ cells. Transposable elements can double single genes or groups of genes; they can reorganize genes by combining them differently, and they can change the local position of genes (the reason they are sometimes also called jumping genes) and thus bring genes under the control of other genetic switches (Pennisi 2007). These phenomena can be interpreted as a form of "self-organization" of the genome, but they also include the possibility of the creation of new genes. The activity of the transposable elements is under the control of the cell, which, according to some biologists, is considered to be a cognitive entity. The cell in turn is under the impact of ecological conditions. The cell has several possibilities to exert an impact on the activity of transposable elements. It can create new transposable elements, or can control and even restrict their activity. In the latter case, the deactivation of transposable elements—a mechanism that is called RNA-interference because it blocks or destroys RNA molecules—exerts stability in the phenotype of an organism when there is no drastic change in the ecological conditions. However, if, by contrast, radical and lasting ecological changes prevail, the organism gives information to the cells which—through the weakening of the RNA interference and easing of the control of transposable elements—in turn induces a creative process of reorganization of the genome. From these findings, some biologists have concluded that changes to the genome are not mainly a consequence of random point mutations, but are due to reactions of the cell, which responds to radical and stressful changes in the ecological conditions, in the form of nuclear radiation, contact with varmints, extreme diet, water scarcity, injuries, etc. (Bauer 2008, pp. 23–30, 84–94).

One of the most surprising new findings in this context was the discovery that changes in the environment can exercise a permanent influence on the gene regulation and gene expression, and through this mechanism heritable changes in the phenotype can be induced, e.g. in the form of RNA-interference (Bauer 2008, p. 26). Epigenetics is the study of heritable changes in the phenotype or gene expression caused by non-genetic mechanisms. In a wider sense, it refers to all mechanisms other than changes in the underlying DNA-sequence that influence the embryonic development. Epigenetics can hence be interpreted as a (counter-) reaction to the dominance of genetics in the modern synthesis. Typically, changes in the environment induce chemical modifications of the DNA, which turn a gene on or off. This process does not represent a mutation; rather, it exerts effects similar to a mutation and hence constitutes an epigenetic mark that can be passed on to the next generation. Epigenetic inheritance can lead to changes in the phenotype that can last for generations. For example, it has been found that extreme diet during gestation can change epigenetic patterns of the descendants and new traits can arise that can last for generations and also affect fertility (Pennisi 2008, p. 197). Thus, environmental conditions influence the heritable phenotype (Jablonka and Lamb 2005, p. 143). At first glance, epigenetic inheritance seems to question one of the dogmas of the modern synthesis, the so called central dogma of molecular biology (articulated by Francis Crick in 1958) according to which information moves only from nucleic acids (DNA and RNA) to proteins, and not vice versa. But since epigenetic inheritance is more about the impact of RNA on DNA, it is said not to contradict the central dogma (Jablonka and Lamb 2005, p. 152). However, the findings according to which epigenetic variations in somatic cells can indeed induce changes in germ cells (Jablonka and Lamb 2005, p. 149) are evidence for the breakthrough of the Weismann barrier. In addition, the theory of epigenetic inheritance has posed the question of a rehabilitation of Lamarckism, since the theory postulates that changes in phenotypic variants are inherited (Jablonka and Lamb 2005, p. 143). Contrary to popular interpretations, this view is mostly rejected by most of the more unorthodox evolutionary biologists. They mostly argue that Lamarck explained the inheritance of the acquired characteristics differently, namely through physiological adaptation. Though epigenetic inheritance has been confirmed in countless experiments, whether it represents an evolutionary principle is still an open question. Some doubt it, because epigenetic traits are gradually lost over a few generations and thus are reversible. There is also no sound empirical evidence that epigenetic variants are adaptive (Jablonka and Lamb 2005, p. 153).

These discoveries in molecular biology have also led to a reassessment of mutations, which are unevenly distributed, as an analysis of the genome of several species has shown. It seems that the cell has the capacity to allow only some parts of the genome to be subject to mutations, mainly those parts that have been duplicated by transposable elements, whereas Hox-genes and other important sequences are less subject to mutations. In other words, the original Hox-genes have been retained, while the copied genes have been released for change, e.g. for the reorganizations of mutations. The conclusion to be drawn here is that mutations are not completely random and that the cell shows a high capacity to control the genomic architecture (Bauer 2008, p. 66, pp. 121–125, p. 135).

Furthermore, it has been widely recognized that, in the early stages of evolution and also in more recent times, not only gene transfer from one generation to the next (vertical gene transfer) took place but also massive horizontal gene transfer, i.e. gene exchanges between individuals of completely different species. Transposable elements (transposons) cannot merely change the structure of a genome within an organism (see above), but can also transfer genes from the genome of one organism into the genome of another organism (Bauer 2008, p. 27). This phenomenon was an important creative factor of recombination. For example, mammals took over genes from viruses, and today it is known that more than 220 genes in the human genome are due to horizontal gene transfer (Bauer 2008, p. 49). The empirical evidence of horizontal gene transfer also implies that phylogenetic trees have to be reconsidered in the sense that they no longer have to be presented as linear lines that branch out, but can be conceived as networks. On this view, evolution is no longer considered to be linear but may be viewed as reticulate.

Another form of inheritance, called "ecological inheritance", has caused attention. It is also a phenomenon that brings ecosystem ecology and evolutionary biology together. Ecological inheritance is due to niche construction, that is, the construction of the environment out of elements of the external world by individual organisms. In this way, they alter the selection pressure of the environment. Ecological inheritance occurs when transformed habitats—by niche construction modified selection pressures—are transmitted to the descendants. Niche construction is considered to be incompatible with the so called external view of the modern synthesis, according to which the internal properties of organisms are only determined by characteristics of the external environment. By contrast, the niche construction theory postulates that organisms cause changes in their environment, which makes niche construction not only a consequence of the evolution, but a discrete factor of evolution (Odling-Smee 2010, pp. 176–180).

In another research branch, a multitude of findings in the last decades suggest that co-operation is an important principle of evolution, in particular, in the creation of new species. For example, in early evolution, eukaryotic cells came into existence because archaea-cells imported bacteria and let them become part of their cells, the so called endosymbiosis (Margulis 1970, 1981). This important evolutionary step was a precondition for the evolution of animals and humans and cannot be explained with the principles of the modern synthesis, but only with co-operation (Bauer 2008, pp. 52).

Since phenomena such as endosymbiosis, horizontal gene transfer, and gene duplication, recombination of genes, and changes in the regulation of genes have played a considerable role in the creation of variety, it is argued that the role of point mutations in the creation of new species has been overestimated by the modern synthesis. Recent discoveries suggest that speciation has primarily arisen from radical changes in the genome, which in turn was mainly induced by eruptive activations of the transposable elements (Bauer 2008, p. 56, p. 84). Furthermore, extensive paleontological studies provide more and more evidence that the incidence of species disappearance is mainly due to punctual mass extinctions, such as volcanism, mega-ice ages, etc. Consequently, natural selection appears not to be the predominant cause either of speciation or of species disappearance (Bauer 2008, p. 100). Whereas the randomness on the micro-level is being more and more questioned (see the discussion of mutations above), on the macro-level, randomness seems to be still an evolutionary factor.

3 Recent Developments in Evolutionary Biology: Assessment

In this section, the question is addressed as to the way in which biologists have assessed the recent findings, before turning to its relevance for economics in Sect. 4.

The research program of evo-devo has led to a complete change in the view of the manner of the evolution of complex biological systems. Already in 1994 Müller wrote: “Of course, Neo-Darwinist theory has in its core statements been confirmed, and generally been accepted” (Müller 1994, p. 160).[3] But then: “The inclusion of internal ‘organismic’ factors would not only mean an extension of the synthetic theory but would amount to a paradigm shift” (Müller 1994, p. 186).[4] Müller argues that, in the modern synthesis, the only forces that give evolution a direction are external factors, whereas under a system theory of evolution, variations would be

[3] Translation from German by the author of this paper.

[4] Translation from German by the author of this paper.

considered to be dependent on internal factors, namely the path-dependent rules of embryonic development (Müller 1994, p. 186).

Another representative of evo-devo, Carroll, says: “First, I assert that Evo Devo constitutes the third major act in a continuing evolutionary synthesis. Evo Devo has not just provided a critical missing piece of the modern synthesis—embryology—and integrated it with molecular genetics and traditional elements such as paleontology. The wholly unexpected nature of some of its key discoveries and the unprecedented quality and depth of evidence it has provided toward settling previously unresolved questions bestow it with a revolutionary character” (Carroll 2006, p. 283). However, Carroll does not see any contradiction with the main principles of the modern synthesis. In the last chapter of his book on evo-devo, he adheres to the traditional view of evolution as a combination of “completely random ...[ ]... genetic variation(s) by mutation” as well as “nonrandom” and “powerful” sorting or selection proceses (Carroll 2006, p. 290).[5] As Müller puts it: “Some regard evo-devo as perfectly compatible with a strictly selectionist theory of evolution; others claim it represents a strong departure from it” (Müller 2005, p. 88).

Kirschner and Gerhard see their theory of facilitated variation as a supplement to and completion of Darwin’s theory (Kirschner and Gerhart 2007, p. 304, p. 319). They also raise the question of whether their theory can contribute to an understanding of social, political or technical elements. Subsequent to some careful suggestions in the form of analogies, they conclude that at least their proposition provokes metaphors other than those of Social Darwinism. They suggest that history (in general) is not only the product of selection determined by external factors and competition, but also the structure and capacity of societies and organizations to evolve, to adapt and to renew (Kirschner and Gerhart 2007, p. 357).

Prominent representatives of epigenetic inheritance, Jablonka and Lamb, write in the foreword of their influential book “Evolution in Four Dimensions”: “Our basic claim is that biological thinking about heredity and evolution is undergoing a revolutionary change. What is emerging is a new synthesis, which challenges the gene-centered version of neo-Darwinism that has dominated biological thought for the last fifty years.” (Jablonka and Lamb 2005, p. 1) Their main critique is thus not the focus of the modern synthesis on natural selection, but its gene-centrism mainly in heredity. They state that “evolutionary change can result from instruction as well as selection”, the meaning of instruction being “internal or external regulatory signals” that control the activity in cells and organisms (Jablonka and Lamb 2005, p. 102). In their study, they not only discuss the genetic inheritance system, but also the epigenetic, behavioral and symbolic ones. In doing so, they suggest classifications of the four systems according to their predominant hereditary transmission, unit and origin of variation, target of selection, unit of evolution,

[5] “... the sorting of these variations as to which will persist and which will be discarded is determined by a powerful, selective and nonrandom process “(Carroll 2006, p. 290).

reproduction of information, transmitting of information, and so on (Jablonka and Lamb 2005, p. 39, p. 234, p. 236).

What makes an overall assessment difficult is the fact that the same discovery leads exponents of the modern synthesis and radical critics to converse interpretations. For example, in a standard textbook on evolution, one can still read “The genes of most TEs [transposable elements] do not contribute to development or function of the host organism; rather, they encode only proteins essential for replication and transposition of the retroelement itself. They are an example of a selfish genetic element, or ‘selfish gene’”(Futuyma 2007, p. 459). This is in contrast to Bauer, who states that genes cannot be interpreted as egoistic since they are under the control of the cell and not autonomous (Bauer 2008, p. 37). Being familiar with recent discoveries concerning transposable elements, Bauer claims: “Genes and respectively genomes follow three basic biological principles [ ] : co-operation, communication and creativity” (Bauer 2008, p. 17).[6] In Bauer’s understanding of evolution, selection plays a role, but only in the sense of a tautology: Individual organisms that are nonviable cannot survive and those that are not propagable will not reproduce (Bauer 2008, p. 104, p. 188). In an introduction to the extended synthesis, Pigliucci and Müller state: “On this view, natural selection becomes a constantly operating background condition, but the specificity of its phenotypic outcome is provided by the developmental systems it operates on. Hence the organisms themselves represent the determinants of selectable variation and innovation” (Pigliucci and Müller 2010b, p. 13).

Though it is definitely too early to give a conclusive appraisal of the recent findings in evolutionary biology, the least one can say is that a thorough reassessment of the significance of chance, genes and selection in biological evolution is inevitable. From an evolutionary economics viewpoint, the most interesting question concerns the role of selection as an evolutionary principle, since this is the one that has been the most transferred to economics and the least questioned in economics. Most of the representatives of the recent research findings presented here (Jablonka and Lamb 2005 are an exception), as well as those of the critical approaches before 1990 (as hinted at in the introduction), agree that the predominance of natural selection is one of the weakest elements of the modern synthesis. As mentioned above, the modern synthesis has acknowledged the phenomenon of genetic drift (Senglaub 1998, p. 574), which implies that, even within the modern synthesis, the statement “no evolution without selection” cannot be maintained.[7] By contrast, in the extended synthesis, not primarily due to genetic

[6] Translation from German by the author of this paper.

[7] Genetic drift is the change in the relative frequency with which a gene variant occurs in a population when changes in the frequency of gene variants occur randomly. Genetic drift is acknowledged as an evolutionary principle, since it can reduce genetic variability. Changes in the gene pool are then considered to be either the consequence of random fluctuations in proportions (gene drift) or of non-random adaptations due to better adaptation (natural selection) (Futuyma 2007, p. 10). Depew/Weber’s interpretation in 1994 was: “If this theory [neutral theory by Kimura] is generally true, natural selection can no longer be presumed to be even heuristically

drift but to other discoveries such as the importance of internal regulation or of niche construction, the significance of natural selection as a factor of evolution is not completely denied but definitely pushed back. Instead, internal factors and regulatory changes (also in the creation of new species) are emphasized.

It can be concluded that the recent findings in biology outside the mainstream suggest a shift away from random mutations, competition, and selection, towards co-operation, internal regulation, networks, self-organization, and path-dependency. In particular and contrary to the modern synthesis, some findings in evolutionary biology suggest that (amongst others) (1) mutations are not completely random and variety is not only triggered by mutations, (2) selection is not the predominant evolutionary factor in the modification of species, the origin of new species and the extinction of species, (3) the systemic character of the genome, including phenomena of co-operation, self-organization and communication, has been underestimated (also in the generation of novelty), and (4) there is more and more evidence on discontinuous evolution.

The presented research programs in biology have in common the fact that their representatives do not reject the modern synthesis completely, all of them being convinced of evolution and the principle of common descent.[8] It should also be stressed that the research findings presented above are generally confirmed and accepted. Many of these findings and theories are not inconsistent with the modern synthesis, but that does not necessarily imply that they are fully covered by the modern synthesis framework (Pigliucci and Müller 2010b, p. 4). Thus, the overall assessment and interpretation with respect to the modern synthesis is still quite controversial. Some say they are still satisfied with the modern synthesis, others suggest an extension of the modern synthesis[9] or even a new evolutionary synthesis and “some say it’s time for modern synthesis 2.0” (Pennisi 2008, p. 196). Then there are those who have declared “the end of Darwinism” (Bauer 2008).[10]

the primary agent in evolutionary processes, and genetic drift, or something like it, can no longer be blithely treated as a trivial or merely annoying secondary evolutionary force” (Depew and Weber 1994, p. 363). Though the empirical evidence of genetic drift is now unambiguous, the question of the relative importance of selection and genetic drift still prevails in evolutionary biology today and is another instance of a possible completely divergent interpretation of the significance of an empirical phenomenon. (For a discussion of the different perspectives, see Kutschera and Niklas 2004, p. 269). For example, according to new findings, some exponents of the modern synthesis argue that genetic drift can no longer be considered to be crucial for speciation and they conclude that “[n]obody really doubts that most of the body parts have been formed by selection” (Orr 2009, p. 15). This demonstrates that, though genetic drift is recognized in the modern synthesis, it is degraded to a side effect or exception.

[8] En passant it shall be noted that the recent findings in evolutionary biology—e.g. the finding of the ancient origin of the genes for building all kinds of animals—confirm the principle of common descent (Carroll 2006).

[9] In some textbooks on biological evolution, the extended synthetic theory is already explicitly mentioned as the successor of the Synthetic Theory (Kutschera 2008, p. 83).

[10] In the latest edition of his book, Bauer (2010) substituted the subtitle “evolution as a creative process” for “the end of Darwinism”, intending to avoid any confusion about his belief in the fact of evolution. Indeed, he declares the fact of evolution to be “irrevocable” (Bauer 2010, p. 9).

Actually, the concept of a new framework, called the "extended synthesis", seems to be gaining ground. Though the term indicates that the modern synthesis has only to be expanded (to embryology and developmental biology and ecology) yet will remain untouched in its core principles, the claims of the exponents of an extended synthesis are much higher: According to their conception, evolutionary biology also has to carry out a shift in the importance of evolutionary factors. From a methodological point of view, they postulate a shift away from statistical correlation analysis (common in population dynamics) to mechanistic causation research, e.g. "a theory of the mechanistic conditions for the origin and innovation of traits." (Pigliucci and Müller 2010b, p. 13) These are important aspects because some economists seem to have the conviction that whatever new discoveries in biology might emerge, the main explanation of evolution will always center on mutations, genes and selection, everything else being extensions, side-effects or exceptions. To take the recent developments in evolutionary biology seriously means to renounce this false conception.

The variety of theoretical approaches in biology can also be viewed in a broader context. In the philosophy of science, it is argued that, since 1970, in science in general, and not only in biology, there has been a tendency from a monolithic towards a pluralistic understanding of science. In biology, this would imply that the modern synthesis will not be substituted for but be supplemented by a range of distinct alternative approaches (Beurton 1995, p. 119).[11] This would be (or already is), by the way, a very similar development to that of economics.

4 Consequences for Evolutionary Economics

We have seen that, at present, there is a pluralism of scientific interpretations of evolution: Traditional interpretations within the modern synthesis go along with new ones that explicitly break with the modern synthesis without denying common descent and without completely rejecting the role of random genetic modification and selection. This pluralism cannot be denied. What makes an assessment for economists really difficult is that those biologists who have come up with the new concepts and principles do not agree on the significance of these new findings and their relation to the mainstream paradigm, the modern synthesis.

One reason why biological scientists draw different conclusions from the new findings is not so much grounded in professional controversies as it is due to the implications of the research results for the social and cultural life in the human societies in which we actually live. For example, Carroll's motivation to adhere to the random variation-selection-paradigm is—according to my interpretation—his aim to support teaching in biological evolution based on scientific results and to provide a multitude of arguments against creationism and intelligent design (Carroll

[11] Quoted from Senglaub (1998, p. 577).

2006). Considering the anti-Darwinian and anti-evolutionary movements of creationism and intelligent design, my impression is that it seems to be too difficult for scientists—mainly in the U.S.—to make clear to the public that, in the scientific community, there is unanimity that evolution has definitely taken place but a controversy on how it has happened. In a non-scientific public discussion, a scientific discipline (evolutionary biology) that is not united in the explanation of its main research object (evolution) is much weaker than one that adheres to common principles. In addition, critics of the modern synthesis risk—not only in the U.S.—being expelled from the scientific community (Bauer 2008, p. 110). Consequently, most of the biologists in the profession strive to avoid this dispute. This is not the case for Bauer, whose goal is to stress the role of co-operation and communication in evolution in order to demonstrate that the idea of egoism as a central evolutionary principle as well as the consideration of living beings as machines is not based on solid scientific research. He argues that this is important to acknowledge since concepts such as those of egoistic genes generate anthropological models that have negative consequences for the way we actually live in human societies (Bauer 2008).

Although not all scholars of evolutionary economics are inclined to resort to theories in biology, there are two arguments as to why these recent findings in evolutionary biology should be of interest to the scientific community of evolutionary economists: (1) For those evolutionary economists who have always found that the fact of evolution in general is more inspiring than the theory, new findings in evolutionary biology still will have effects but only indirect ones. Theories always influence mental attitudes and the way in which studies are conducted and empirical results are looked at. As Zimmerli put it: “Each scientific theory sediments into human consciousness in a way that its theorems are eventually considered to be elements of reality” (Zimmerli 1990, p. 138). (2) In biology, critics of the modern synthesis argue that (mainstream) biology is very much shaped by mechanical physics and economics, and that, in general, beings are considered to be machines and not living systems (Bauer 2008, p. 13). Thus, the danger in evolutionary economics is that, in resorting to the modern synthesis, indirectly all the mechanical aspects this research area has intended to avoid are re-imported. One of the common characteristics of biology and economics—in contrast to physics—is that they deal with living organisms. Consequently, in evolutionary economics, one of the crucial questions should be: What are the main and specific properties of living systems[12] and, in particular, human beings? The mechanistic models in parts of the extended synthesis unfortunately are neither very helpful in answering these questions nor in providing fruitful heuristic devices towards these questions.

In the past, the influences of theories of evolutionary biology on economics were manifold and can be summed up by different heuristic strategies (Witt 2004, 2008), two of which will be dwelled on further in the following.

[12] It is not the subject of this paper to discuss this question, but important characteristics are sentiments and perceptions. See also Capra (1996, part IV).

First, an important way in which biology has had an influence on economics is by the methodological device of analogies. It is widely acknowledged that analogies in sciences can have important heuristic functions and they can be used as a method to build new hypotheses. It should be emphasized that the method of analogy as a heuristic device in science (Klamer and Leonard 1994, p. 35) as such cannot be criticized. In particular, it cannot be argued that the analogy is inadequate because some characteristics, elements or relations from the original theory (in one science) do not correspond with those of the destination theory (in another science), because it is the essence of the analogy that not all aspects match. Thus, the statement that an analogy is inadequate or useful, or more general, good or bad, is illegitimate since the analogy is a method and not the correspondent, counterpart or equivalent itself. The problem with these kinds of analogies is that they might become a justification in the sense that, without further arguments, studies or proofs, the findings from one science are directly transferred to another science (Ruse 1986, p. 33). Analogies hence run the risk of starting a life of their own without being reassessed and verified. This holds in particular for analogies based on the variation-retention-selection paradigm. Theories based on these analogies often fail to inquire as to whether the principles of variation, retention and selection are indeed the core principles of evolution in economic reality and whether omitting specific economic, social and cultural factors is a proper approach.

Reflecting on biological evolutionary theorizing in the last 50 years, the impression arises that a lot of analogies are based more on the original Darwinism of the nineteenth century or the early modern synthesis of the 1930s and 1940s. This is astonishing even in light of the state of evolutionary biology in the second half of the twentieth century. Some, thus, have wondered how economists could, on the one hand, criticize neoclassical economics and, on the other hand, prefer competition to self-organization (Foster 2001, p. 123).[13] The concentration on selection analogies would not have been a major problem if analogies had only the function of inspiration and would not become independent. In light of the recent developments in evolutionary biology, is it then advisable to build analogies in economics based on evo-devo, plasticity, epigenetic inheritance and niche construction? As hinted above, the main function of metaphors and analogies is to create hypotheses and, in fulfilling this function, they cannot be preferred to any other. New hypotheses in biology do not necessarily provide better hypotheses in economics than old ones, since biology and economics are not congruent anyway.

Second, since the explanation of biological evolution is undergoing a reassessment, in which the traditional core principles of evolution, such as blind mutation and selection, lose significance (although as argued above, they do not become completely irrelevant), concepts such as generalized Darwinism (Aldrich et al. 2008) are probably no longer up to date. As has already been argued by Cordes (2006, p. 532), generalized Darwinism does not even take into account all those

[13] Exceptions were e.g. Silverberg et al. (1988), Foster (1997), Witt (1997).

principles that mainstream biologists consider to be essential to Darwin's theory. For example, the principles of descent and of speciation are not chosen to belong to generalized Darwinism. Buenstorf (2006, p. 515) states that, since generalized Darwinism turned the heuristic frame into an ontological fact, subjects that are not covered by the assumed three core principles of evolution are excluded from examination, e.g. certain forms of learning and knowledge transfer. In this paper, I further argue that generalized Darwinism does not take into account the principles of co-operation and self-organization. Though the representatives of generalized Darwinism concede that to explain evolution sufficiently "auxiliary" principles are required (Aldrich et al. 2008, p. 592), the question that arises is why co-operation and self-organization are not considered as core principles of generalized Darwinism and what the selection criteria are that make the difference between core and auxiliary principles of evolution.

As mentioned above, not all evolutionary economists are inclined to let themselves be inspired by theories of biological evolution. In particular, analogies and concepts such as generalized Darwinism bear the danger of neglecting specific human, social and economic aspects of evolution and, in addition, their explanative power is rather humble. Events other than economic crises should remind us that there is still much to learn about human behavior and economic systems. Thus, instead of looking for common principles with other scientific disciplines, I would argue in favor of concentrating more on the specific characteristics of human, social and economic evolution. This kind of analysis can acknowledge that human beings have animal bodies and hence are to a certain extent subject to biological evolution. However, in addition, the evolution of societies and economies follows certain rules and norms, some of which are, at least in part, the result of deliberate decisions and not subject to biological evolution. The continuity hypothesis (Witt 2004, pp. 129–133) is, for example, a heuristic strategy which is in line with these requirements.

5 Conclusion

In this paper, it has been shown that, in addition to modern synthesis principles such as random mutations and selection, the principles of co-operation and self-organization are considered by quite a number of researchers in biology today to be core principles of biological evolution. With respect to the identification of important factors of evolution, it must also be taken into consideration that the scientific development of evolutionary biology will continue. The recent discoveries can be interpreted as an indication of the deep changes (maybe including paradigm shifts) that this science might still undergo in the next decades.

The short review above hopefully has documented that nobody in biology really doubts the existence of mutations and natural selection. But some researchers question their assumed significance which is magnified and stylized to a transdisciplinary variation-retention-selection paradigm. The problem with this paradigm is that it ignores certain important aspects, in particular, internal regulation, co-operation and self-organization.

As indicated in the introduction, in the second half of the twentieth century, the modern synthesis had already shown quite a remarkable resistance towards a wave of critique. In our own area of expertise, we have experienced that mainstream paradigms are highly innovative and adaptable in order to insure their survival. A similar development can take place in evolutionary biology and the modern synthesis might absorb the recent findings. Fortunately, evolution—also the one of sciences—can lead to surprises. Though it seems as if genes and natural selection will continue to be considered as core factors of evolution, the focus of attention can also shift to other evolutionary principles. How things develop is not only dependent on the critical reception in biology, but also on debates in other sciences as in economics.

References

Aldrich HE et al (2008) In defence of generalized Darwinism. J Evol Econ 18:577–596

Bauer J (2008) Das kooperative Gen – Abschied vom Darwinismus. Hoffmann und Campe, Hamburg

Bauer J (2010) Das kooperative Gen – Evolution als kreativer Prozess. Heyne, München

Beurton PJ (1995) Neo-Darwinism or synthesis? In: Wolters G, Lennox JG (eds), Concepts, theories, and rationality in the biological sciences. Universitäts-Verlag Konstanz/Pittsburgh Press, Konstanz/Pittsburgh, PA

Buenstorf G (2006) How useful is generalized Darwinism as a framework to study competition and industrial evolution? J Evol Econ 16:511–527

Capra F (1996) The web of life: a new synthesis of mind and matter. HarperCollins, New York, NY

Carroll SB (2006) Endless forms most beautiful. The new science of evo devo and the making of animal kingdom. Weidenfeld & Nicolson, London

Cordes C (2006) Darwinism in economics: from analogy to continuity. J Evol Econ 16:529–541

Depew DJ, Weber BH (1994) Darwinism evolving: system dynamics and the genealogy of natural selection. MIT, Cambridge, MA

Eldredge N (1985) Unfinished synthesis: biological hierarchies and modern evolutionary thought. OUP, New York, NY

Foster J (1997) The analytical foundations of evolutionary economics: from biological analogy to economic self-organization. Struct Change Econ Dynam 8:427–451

Foster J (2001) Competition, competitive selection and economic evolution. In: Garrouste P, Ioannides S (eds) Evolution and path dependence in economic ideas. Edward Elgar, Cheltenham, pp 107–132

Futuyma DJ (2007) Evolution. Elsevier, Munich

Gilbert SF, Epel D (2009) Ecological developmental biology. Integrating epigenetics, medicine, and evolution. Sinauer, Sunderland, MA

Gould SJ (1980) Is a new and general theory of evolution emerging? Paleobiology 6:119–130

Gould SJ, Eldredge N (1972) Punctuated equilibria: an alternative to phyletic gradualism. In: Schopt TJM (ed) Models in paleobiology. Freeman, Cooper, San Francisco, pp 82–115

Gould SJ, Eldredge N (1977) Punctuated equilibria: the tempo and mode of evolution reconsidered. Paleobiology 3:115–151
Jablonka E, Lamb MJ (2005) Evolution in four dimensions. Genetic, epigentic, behavioral, and symbolic variation in the history of life. MIT, Cambridge, MA
Kauffman SA (1993) Origins of order: self-organization and selection in evolution. Oxford University Press, Oxford
Kimura M (1955) Solution of a process of random genetic drift with a contiuous model. Proc Natl Acad Sci USA 41:144–150
Kimura M (1983) The neutral theory of molecular evolution. CUP, Cambridge, MA
Kirschner MW, Gerhart JC (2005) The plausability of life. Yale University Press, New Haven
Kirschner MW, Gerhart JC (2007) Die Lösung von Darwins Dilemma (German). Rowohlt, Reinbek
Klamer A, Leonard TC (1994) So what's an economic metaphor? In: Mirowski P (ed) Natural images in economic thought. CUP, Cambridge, MA
Kutschera U (2008) Evolutionsbiologie, 3rd edn. Eugen Ulmer, Stuttgart
Kutschera U, Niklas KJ (2004) The modern theory of biological evolution: an expanded synthesis. Naturwissenschaften 91:255–276
Lima-de-Faria A (1988) Evolution without selection. Form and function by autoevolution. Elsevier, Amsterdam, The Netherlands
Margulis L (1970) The origin of eukaryotic cells. Yale University Press, Haven
Margulis L (1981) Symbiosis in cell evolution. Life and its environment on the early earth. Freeman and Company, San Francisco, CA
Maynard Smith J (1964) Group selection and kin selection. Nature 201:1145–1147
Maynard Smith J et al (1985) Development constraints and evolution. Q Rev Biol 60:265–287
Müller GB (1994) Evolutionäre Entwicklungsbiologie: Grundlagen einer neuen Synthese. In: Wieser W (ed) Die Evolution der Evolutionstheorie – Von Darwin zur DNA. Spektrum Akademischer Verlag, Heidelberg, pp 155–193
Müller GB (2005) Evolutionary developmental biology. In: Wuketits FM, Ayala FJ (eds) Handbook of evolution. Wiley-VCH, Weinheim, pp 87–115
Müller GB, Newman SA (2003) Origination of organismal form: the forgotten cause in evolutionary theory. In: Müller GB, Newman SA (eds) Origination of organismal form – beyond the gene in developmental and evolutionary biology. MIT, Cambridge, MA, pp 3–10
Odling-Smee J (2010) Niche inheritance. In: Müller GB, Pigliucci M (eds) The extended synthesis. MIT, Cambridge, MA, pp 175–207
Orr HA (2009) Selektion im Test. Spektrum der Wissenschaft Spezial 1:12–19
Pennisi E (2007) Jumping genes hop into the evolutionary limelight. Science 317:894–895
Pennisi E (2008) Modernizing the modern synthesis. Science 321:196–197
Pigliucci M, Müller GB (eds) (2010a) The extended synthesis. MIT, Cambridge, MA
Pigliucci M, Müller GB (2010b) Elements of an extended synthesis. In: Pigliucci M, Müller GB (eds) The extended synthesis. MIT, Cambridge, MA, pp 3–17
Reif WE, Junker T, Hoßfeld U (2000) The synthetic theory of evolution: general problems and the german contribution to the synthesis. Theory Biosci 119:41–91
Riedl R (1975) Die Ordnung des Lebendigen. Parey, Hamburg, Berlin
Ruse M (1986) Taking Darwin seriously. Basil Blackwell, Oxford
Senglaub K (1998) Neue Auseinandersetzungen mit dem Darwinismus. In: Jahn I (ed) Geschichte der Biologie, 3rd edn. Spektrum Akademischer Verlag, Heidelberg, pp 558–579
Silverberg G, Dosi G, Orsenigo L (1988) Innovation, diversity and diffusion: a self-organization model. Econ J 98:1032–1054
Vrba E, Eldredge N (1984) Individuals, hierarchies and processes: towards a more complete evolutionary theory. Paleobiology 10:146–171

Wieser W (1994) Gentheorien und Systemtheorien: Wege und Wandlungen der Evolutionstheorie im 20. Jahrhundert. In: Wieser W (ed) Die Evolution der Evolutionstheorie – Von Darwin zur DNA. Spektrum Akademischer Verlag, Heidelberg, pp 15–48

Witt U (1997) Self-organization and economics – What is new? Struct Change Econ Dynam 8:489–507

Witt U (2004) On the proper interpretation of 'evolution' in economics and its implications for the production theory. J Econ Methodol 11:125–146

Witt U (2008) What is specific about evolutionary economics? J Evol Econ 18:547–575

Zimmerli WC (1990) Grenzen des evolutionären Paradigmas – Metakritische Bemerkungen zu einer auslaufenden Denkmode. In: Jüdes U, Eulefeld G, Kapune T (eds) Evolution der Biosphäre. Hirzel, Stuttgart, pp 137–152