A Dynamic Mind

Perspectives on Industrial Dynamics
in Honour of Staffan Laestadius

Editors: Pär Blomkvist and Petter Johansson
A Dynamic Mind
Perspectives on Industrial Dynamics in Honour of Staffan Laestadius

A 'Vänbok' to professor emeritus Staffan Laestadius
from the Division of Sustainability and Industrial Dynamics (SID),
Department of Industrial Economics and Management (INDEK), KTH.
A Dynamic Mind.

Perspectives on Industrial Dynamics in Honour of Staffan Laestadius.

EDITORS Pär Blomkvist & Petter Johansson.


LAYOUT Susanne Blomkvist.

Published by the Division of Sustainability and Industrial Dynamics, Department of Industrial Economics and Management (INDEK), KTH.

First edition, First print.

Stockholm, November 2016.
Table of Contents

Foreword .................................................................................................................................................. 7

About the book ....................................................................................................................................... 8

1: This is Industrial Dynamics
Pär Blomkvist, Petter Johansson & Staffan Laestadius ........................................................................ 13

2: Eric Dahmén and Industrial Dynamics
Staffan Laestadius .................................................................................................................................. 23

3: Systems thinking in Industrial dynamics
Pär Blomkvist & Petter Johansson ......................................................................................................... 45

Staffan Laestadius & Annika Rickne ...................................................................................................... 75

5: Location as a Matrix of Competition
Cali Nuur .................................................................................................................................................. 109

6: Diffusion of Innovations
Emrah Karakaya & Pranpreya Sriwannawit Lundberg ...................................................................... 151

7: Technological Transformation in Process Industries
Michael Novotny ..................................................................................................................................... 177

8: The Route Towards a Cashless Society in Sweden
Niklas Arvidsson .................................................................................................................................... 203

9: Dynamics: in the ICT Industry
Vicky Long ................................................................................................................................................. 223

10: Structural Change in the Petroleum Industry
Maria Morgunova & Vladimir Kutcherov .......................................................................................... 249

11: Industrial Catching Up
Vicky Long ................................................................................................................................................. 277

12: Policy Instruments for Energy Efficiency
Rurik Holmberg ....................................................................................................................................... 303

13: Dynamics in Swedish Industrial and Political History
Eric Giertz .................................................................................................................................................. 321

Author presentations ............................................................................................................................... 369
Foreword

This is a Vänbok, in English and German, a Festschrift. It is written by members of the division of Sustainability and Industrial Dynamics (SID) at INDEK, KTH. And, most importantly, it is written in honour of Professor Staffan Laestadius. We would like to thank Staffan for his work in building a truly inspiring academic environment and for his role in creating SID. From being quite peripheral, the division has grown steadily and is, thanks to Staffan’s efforts, now an integral part of the department of Industrial Economics and Management (INDEK).

The academic subject is industrial dynamics, which is a research and teaching field based on industrial and technological transformation. It is influenced by both systems and institutional theories and borrows concepts and models from the social sciences (sociology, history, political sciences, business/management, economics, behavioural sciences). However, the focus is on industrial and technological transformations on the meso-level.

In most cases, a Vänbok is written in secret and presented to the unknowing recipient as a surprise. That is not the case for this book. Staffan Laestadius has been highly involved in the whole process and is contributing in several chapters. It seems only appropriate that he should be an integrated part of the book celebrating and presenting his heresy – Industrial Dynamics at KTH.

The idea of a Vänbok was introduced by Pär Blomkvist two years ago and he and Petter Johansson has acted as editors and project leaders. The whole SID division has been involved in this effort to create a common understanding of our academic field and the project has moved through many phases of peer review seminars.

The first motive behind the project was to create a canon in Industrial Dynamics reading. The chapters present some of the most important fields in our area, albeit not a full coverage. The second motive was related to knowledge management and retention. We wanted to create a reservoir of knowledge at a time of transition when some of our most influential members, Staffan included, retired. The third motive was directed towards teaching. The chapters are aimed at master level students at technical universities like KTH.

It is with great joy that we hand over this Vänbok to Staffan Laestadius – a person with a truly Dynamic Mind.

Stockholm November 2016
About the book

This is a joint production by all participating authors, spanning roughly two years. The project has been managed by editors Pär Blomkvist and Petter Johansson. The strategy has been to draw on the existing knowledge among the industrial dynamics oriented researchers at our division and formalising their knowledge.

The book contains thirteen individual chapters. It is divided into three parts (see layout below) with more theoretically focused chapters in the first part, followed by more empirically oriented and case based chapters and in the third part, chapters dealing with broader issues such as policy and institutions. But the chapters are not fully aligned and we have allowed some overlap. Each chapter has its own individual perspective on current research on industrial dynamics and transformation. The texts are not professionally language checked and the authors bear full responsibility for content, language, spelling, references, etc.

The Vänbok is the first phase in a bigger project that we have called the “three stage rocket”:

1. We publish this book as a TRITA Working paper in November 2016 and hand it over to Staffan Laestadius as a Vänbok at a ceremony at INDEK. We invite friends and colleagues from KTH and other places. The ambition is that the book release could be the start of a network of researchers in Industrial Dynamics.

2. The SID-homepage (our knowledge hub) is launched in parallel during 2016/2017. This homepage is to be a resource for teachers at SID when giving courses. It contains all the chapters in the SID-book (and additional chapters added over time). Furthermore, the home page contains videos where the authors in the SID-book explain key concepts. We also hope to get video contributions from other prominent figures in our field. In the beginning the homepage is open only for teachers at SID, but we want to make it public in the future – thus establishing the KTH/INDEK Industrial Dynamics Knowledge Hub.

3. We aim to publish revised versions of the book chapters, and hopefully additional chapters from authors at our division and invited authors from the network, at an international publisher. The preparatory work for this third stage, starts during the spring of 2017.
## Layout

<table>
<thead>
<tr>
<th>Part 1: Theories and concepts</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>1: This is Industrial Dynamics</strong></td>
</tr>
<tr>
<td>Pär Blomkvist, Petter Johansson &amp; Staffan Laestadius</td>
</tr>
<tr>
<td><strong>2: Eric Dahmén and Industrial Dynamics</strong></td>
</tr>
<tr>
<td>Staffan Laestadius</td>
</tr>
<tr>
<td><strong>3: Systems thinking in Industrial dynamics</strong></td>
</tr>
<tr>
<td>Pär Blomkvist &amp; Petter Johansson</td>
</tr>
<tr>
<td><strong>4: A Critical View on the Innovation Systems Approach</strong></td>
</tr>
<tr>
<td>Annika Rickne &amp; Staffan Laestadius</td>
</tr>
<tr>
<td><strong>5: Location as a Matrix of Competition</strong></td>
</tr>
<tr>
<td>Cali Nuur</td>
</tr>
<tr>
<td><strong>6: Diffusion of Innovations</strong></td>
</tr>
<tr>
<td>Emrah Karakaya &amp; Pranpreya Sriwannawit Lundberg</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Part 2: Case studies of Industrial Dynamics</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>7: Technological Transformation in Process Industries</strong></td>
</tr>
<tr>
<td>Michael Novotny</td>
</tr>
<tr>
<td><strong>8: The Route Towards a Cashless Society in Sweden</strong></td>
</tr>
<tr>
<td>Niklas Arvidsson</td>
</tr>
<tr>
<td><strong>9: Dynamics in the ICT Industry</strong></td>
</tr>
<tr>
<td>Vicky Long</td>
</tr>
<tr>
<td><strong>10: Structural Change in the Petroleum Industry</strong></td>
</tr>
<tr>
<td>Maria Morgunova &amp; Vladimir Kuchorov</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Part 3: Policy, Institutions and Industrialization</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>11: Industrial Catching Up</strong></td>
</tr>
<tr>
<td>Vicky Long</td>
</tr>
<tr>
<td><strong>12: Policy Instruments for Energy Efficiency</strong></td>
</tr>
<tr>
<td>Rurik Holmberg</td>
</tr>
<tr>
<td><strong>13: Dynamics in Swedish Industrial and Political History</strong></td>
</tr>
<tr>
<td>Eric Giertz</td>
</tr>
</tbody>
</table>
A Brief Presentation of the Chapters

In the first chapter, Blomkvist, Johansson and Laestadius introduce the field of industrial dynamics at KTH, and discuss how this academic area has evolved by presenting the three most influential origins: Neo-classical economics, Innovation theory and Evolutionary economics.

In Sweden, the work of Eric Dahmén has formed an important foundation for research on industrial dynamics. In the second chapter Laestadius introduce the work by Dahmén and presents his theoretical concepts and how Dahmén's insights can be used to meet current challenges.

In the third chapter Blomkvist and Johansson discusses systems theory and systems thinking. They present two theoretical frameworks: Large Technical Systems (LTS) and the Multi-Level Perspective (MLP).

In the fourth chapter Rickne and Laestadius present the current research on innovation systems (IS) discussing the foundation for the Swedish research on innovation systems in the 1980s and 1990s.

The fifth chapter by Nuur, is an extensive chapter focusing on the spatial aspects of industrial dynamics (location), with examples from small Swedish Gnosjö to California's Silicon Valley.

Another central theoretical field in industrial dynamics steams from Roger's work on diffusion of innovations. In the sixth chapter Karakaya and Sriwannawit Lundberg discusses the theoretical concepts on diffusion and presents two concrete cases that exemplifies how the theoretical perspectives on diffusion can be used.

In the seventh chapter Novotny presents the case of the bio-refineries related to Swedish pulp and paper industry and uses Dahmén's concept of development block – introduced in the second chapter – for the analysis of the biorefining development.

In the eight chapter Arvidsson discusses the current industrial and societal transformation relating to cash in Sweden, with Sweden as a forerunner in the transition towards the (possibly) cash-less society. In his analysis he uses the Multi-Level Perspective, introduced in the third chapter.

ICT dynamics and the global digitalisation process is discussed by Long in the ninth chapter. This transformation leaves no industry untouched due to the generic character of ICT as a general purpose technology in many fields.
In the tenth chapter Morgunova and Kutcherov discuss the petroleum industry. They use the concepts by Dahmén, from chapter two and seven, to analyse the transition of one of the world’s largest industries.

In the eleventh chapter Long introduce the concept of industrial *catching up* by discussing historical cases from the first and second industrial revolution and present day examples of industrial catching up processes in countries such as China and India.

In the twelfth chapter Holmberg discusses institutional settings in the energy sector and the relation between industrial transformation and policy instruments, using the case of energy efficiency.

Policy is also a central part of the thirteenth chapter. Giertz presents a historical overview of industrial dynamics in Sweden, from the middle of the 19th century until today, and touches upon several of the theoretical concepts and empirical cases introduced in previous chapters in this volume.
1: This is Industrial Dynamics
Pär Blomkvist, Petter Johansson & Staffan Laestadius

The chapters in this book present a set of analytical Perspectives on Industrial Dynamics (ID). In this introductory chapter, we give a background on how the subject area is interpreted at our department (SID) at INDEK, KTH. We also discuss how our field has evolved by 

presenting the three most influential origins of Industrial Dynamics: Neo-classical economics, Innovation theory and Evolutionary economics.

We start by a quote from the first chapter, that may serve as a motivation for the book project and for the usefulness of the research area of Industrial Dynamics:

“This is a period when the world economy has to transform fundamentally in order to rapidly reduce the impact of modern industrial activities and life styles on the climate as well as to adapt to the climate change on the way. To a large extent that transformation has to take place within and between dominating firms and industries which during two hundred years of development have become highly dependent on fossil fuels. This transformation will – and must – impact on technologies and industrial capabilities and it will necessitate management activities as well as policy interventions.” (Laestadius in this volume)

The great challenge ahead for industrial scholars is finding the tools for analysing and managing technology and industrial transformation in climate change. Many projects in the industrial dynamics group at INDEK are related to climate change: The automotive industry is transforming but the electric car is facing a “chasm”. New conditions are created for sustainable energy business. The wind power industry is transforming: in industrial structure, in production costs, in acceptance, in development of new forms of ownership. The transformation of pulp & paper to bio refineries is partly driven by the need to substitute for oil. The diffusion of photovoltaics (the most expensive form of sustainable energy) is also – in many of its applications – the most suitable for developing regions; how to manage that?

These are projects on the meso level – i.e. on the level of technological/innovation systems and industrial structures. The areas under investigation are relevant for strategic management decisions, industrial and technology policy and for understanding transformation to sustainability and green growth: Thus, Industrial dynamics, to a large extent, relates to strategic...
management and policy making on a meso level. The theoretical base of our research, as a branch of the discipline industrial economics and management, has strong relations to a variety of other disciplines such as economics, business administration, economic history and history of science & technology, industrial sociology and economic geography. We use quantitative as well as qualitative research methods.

Innovation is the creation of new combinations, and processes of problem-solving activities constitute the building blocks of innovative creativity. Thus, innovation is a cognitive learning process, irreversible and path dependent. From our research perspective we emphasize that management of knowledge production in its systemic context is a complex process that need much more of integrated research connecting knowledge, learning and innovation. When studying technological trajectories and industrial transformation, we bring science and institutional change into the theory of innovation.

Industrial dynamics is thus a truly multi-disciplinary area of knowledge. At KTH it is characterized by a strong technology focus topic: in research and teaching of industrial phenomena, entrepreneurship and their implications on management and policy our ambition is to integrate that with in depth analyses on the role of technology and the mechanisms and knowledge processes behind technological change. Below we discuss the origin and heritage of industrial dynamics from three interdependent perspectives.

The heritage from Economics

Industrial dynamics has one of its origins in the microeconomic theory within the tradition of neoclassical economics. Or rather in the shortcomings of orthodox economics with its strong focus on markets, on prices and on perfect conditions for equilibrium. What is assumed to constitute the perfectly competitive economy – i.e. homogenous products, perfect information, many actors, no externalities, no economies of scale, etc. – are not only rare situations but also conditions which economic actors strive to avoid. What is often classified as market failures in the textbooks in microeconomics is in fact the normal landscape where industrial and technical actors have to navigate. Economic development does not evolve from everybody producing and selling the same thing but from the creativity taking place behind the markets in structures, like firms and universities, which are more complex than just transmitting market signals.

Before we return to those perspectives in the sections on Innovation theory and Evolutionary economics below, we turn to prof. Bo Carlsson – one of the early introducers of the Industrial Dynamics tradition – and who´s perspective on these issues is as follows:
“Industrial Dynamics (ID) is a new and rapidly growing field of research. Its theoretical roots are similar to those of Industrial Organization (IO), but the questions addressed are different. IO is based on equilibrium (static or comparative static) analysis; there is no causal analysis. In ID the emphasis is on dynamics: seeking to identify and understand the reasons why things are as they are. ID focuses on the causes (driving forces) of economic transformation and growth, and on understanding the underlying processes of transformation, not just the outcomes. The transformation is viewed in its wider historical, institutional, technological, social, political, and geographic context. This means that the analysis often has to transcend disciplinary boundaries and involve multiple dimensions and levels. Economic growth can be described at the macro level, but it can never be explained at that level [...] Economic transformation is a matter of experimental creation of a variety of technologies that are confronted with potential buyers in dynamic markets and hierarchies. Economic growth results from the interaction of a variety of actors who create and use technology, including demanding customers. The interaction takes place in an evolving institutional setting.”

According Bo Carlsson there are five broad themes that constitute the basic questions in industrial dynamics:

1. The causes of industrial development and economic growth, including the dynamics and evolution of industries and the role of entrepreneurship
2. The nature of economic activity in the firm and the dynamics of supply, particularly the role of knowledge.
3. How the boundaries and interdependence of firms change over time and contribute to economic transformation.
4. Technological change and its institutional framework, especially systems of innovation.
5. The role of public policy in facilitating adjustment of the economy to changing circumstances at both micro and macro levels.

Research and teaching at the division of sustainability and industrial dynamics (SID) at KTH touch upon all these five themes. Many SID publications, however, have a strong focus on the role of knowledge and technological activities in systems, industries and firms. To analyse these knowledge formation processes in technology, the technological trajectories emerging and their contribution to changing conditions for economic and

---

social actors it is necessary to go beyond the understanding of prices and markets. This is also what Industrial Dynamics is about.

The heritage from Innovation theory

Innovation theory is central to the subject area. In short, innovation theory has its origin in the tradition emerging from the economist Joseph Schumpeter’s research around the two world wars of the 20th century. It was Schumpeter who introduced the concepts of “innovation” and “entrepreneur” as he was the first to point out the need for venture capital. In the beginning of his research the entrepreneurial problem was in focus; later he studied whether the economy, which over time became more dominated by large corporations could preserve its creativity and capacity for innovation.

The modern innovation research – with roots in the 1970s and 1980s – has largely inherited those old research areas. Not the least Christopher Freeman's *The Economics of Industrial Innovation* (1974) provided inspiration to many innovation researchers. The management aspects of innovation – which also are core issues in the research and teaching at SID/KTH, have their origin in writings of Burns and Stalker (1961) although James Utterback’s popular *Mastering the Dynamics of Innovation* (1994), synthesizing much of innovation research from the late 20th century, may be looked upon as an important milestone in the field.

Significant parts of the management oriented innovation research have focused on identifying and classifying the variety of innovations as regards their technology and function in the industrial system in order to understand their challenges to management and policy. Among all papers published in this area three are frequently quoted:

In *The Patterns of Industrial Innovation* (1978) William Abernathy & James Utterback develop the distinction between radical and incremental innovations as well as product and process innovations and analyse how they change in character when industries/firms mature.

---

A second important contribution in this field, can be found in Henderson and Clark’s distinction between *Modular* and *Architectural* innovation. If a company changes the architecture of a product without changing the core design an architectural innovation is created. This kind of innovation is often triggered by a small modification of a component. Although the alteration of the architecture may not seem as a large alteration it can have a massive impact on the company and result in a leap not unlike the case of radical innovation. One example is the ceiling fan, if a company decides to take the next step and introduce small portable fans it would be an architectural innovation as it involves assembling the components differently. The new product is different from the previous and the acceptance by existing customers might be hard to predict. This new product could be so different from previous products that it requires different sales channels and distribution, possibly attracting new customers. The danger lies in not realizing that an architectural innovation is in fact not incremental.⁸

Clayton Christensen’s *The Innovator’s Dilemma* (1997) provides a third contribution to this research area with his dichotomy between *sustaining* and *disruptive* innovations.⁹ In short sustaining innovations are those that fall in line with existing capabilities and thought styles of a firm and can, thus, even if being radical, be handled by the existing structure of management and engineering. Disruptive innovations, on the other hand, are those that fall outside the fundamental skills and experiences of a firm making it more or less impossible for the firm to adapt to the new technology.

The original Schumpeterian concept of innovation covered all creative combinations that were introduced on the market or in the industrial/economic system. Innovations could thus be organizational as well as technical, related to markets as well as institutions etc. However, for a long time, innovation research has been dominated by a more or less explicit focus on classical mechanic industry. This is now changing.

IKEA’s combination of logistics, design and production with customer involvement is thus entirely in Schumpeter’s spirit as well as its development of new business concepts. Creativity, design and new business models in the digital economy of today also contribute with new aspects of innovative behaviour.

---


This “design discussion” also reveals that innovations are not identical with, or necessarily the result of, science. And innovation theory cannot limit its analytical scope to R&D processes in universities and industrial research labs. The role of science in innovation processes is an open question and a research task for the innovation analyst. Some industries are more R&D based than others. The pulp & paper industry e.g. which has been studied frequently at our division, is historically an industry with low R&D intensity (see Novotny 2016 – this volume). For the ICT industry, of course, the situation is different (see Long 2016 – this volume, on ICT dynamics).

The analyses of innovation processes have a significant impact on our understanding of firms. Firms that learn to innovate can develop capabilities to avoid the price trap of competitive equilibrium which dominates neoclassical analysis. The new theory of the firm, which focuses on how companies may allocate and develop their resources and create a climate contributing to dynamic capabilities is a key starting point in modern enterprise-based analysis of industrial change. The companies' differences as regards capabilities create variety when it comes to competitiveness: also seemingly low-tech businesses can creatively build up competencies that become competitive in global scale. This is basically what successful entrepreneurship and management of innovation and technology is about.

This leads us to the context or embeddedness of industrial and technical change. This embeddedness can be understood – and analysed – from a territorial perspective. We may call this geographical or physical proximity. But context can also be analysed from a cognitive perspective, i.e. proximity of thought styles. This may include engineering communities and cultures more or less related to various industries and technologies and not necessarily localized to a specific region. Embeddedness is also a question of culture, institutions and physical infrastructure. There are several research traditions in this area with strong family resemblances. Some are focused on clusters, others on innovation systems: national and regional, among others (see Laestadius and Rickne in this volume).

At SID/KTH the focus is also on technological systems and their transformation, a transformation where the incentives sometimes come out of necessities, sometimes out of opportunities. Especially in these times of globalization and the emergence of India and China as industrial giants. How to establish and maintain innovative activities and production at the global level? There has also been a focus on development blocks, a Schumpeterian inspired concept developed by Erik Dahmén (See Laestadius in this volume). Neither of these concepts are necessarily restricted to certain territories. The embeddedness, relations and forces may well be characterized by non-territorial proximity.
The heritage from Evolutionary Economics

Industrial Dynamics also has its roots in *Evolutionary Economics*. Although this tradition of economic theory has its origin in the analyses of Alfred Marshall, Thorstein Veblen and, not the least, Joseph Schumpeter the modern approach to non-equilibrium and evolutionary processes of industrial change came with Nelson & Winter’s *An evolutionary theory of economic change* (1982).\(^{10}\)

In short, adopting an evolutionary approach in the analysis makes it possible to get rid of most of the conditions framing the conditions for competitive equilibrium and to focus on paths of industrial and technical transformations. The heritage from evolutionary theory also means that *history matters*. In Industrial dynamics the time line is very important. We argue that earlier choices by actors and historically acquired inertia (sunk costs, technical standards, already built infrastructure, etc.) strongly influences the possibilities for present actors to engage in transformation of technology, technological systems and industry. This focus on historical (and institutional heritage) is mirrored in the use of concepts like these: *learning, structural tensions, co-evolution, technological paradigms/trajectories/regimes, path dependence, product cycles.*

Fundamental concepts in this analytical approach, and imported into economic theory by Nelson & Winter are: *Variation, Selection environment, Adaption, Retention.*

“*Variation* is the creation of a novel technical or institutional form within a population under investigation. *Selection* occurs principally through competition among the alternative novel forms that exists, and actors in the environment *select* those forms which optimize or are best suited to the resource base of an environmental *niche*. *Retention* involves the forces (including inertia and persistence) that perpetuate and maintain certain technical and institutional forms that were selected in the past”\(^{11}\)

In a case study on the dynamic behind the introduction of the moped in Sweden, these evolutionary inspired concepts were used like this:

“In our case the *niche* is the space between the bicycle and the motorcycle as defined by the government in 1952 – the perceived need for a bicycle with a help motor. The *variation* is all the different moped types

---


launched at the market by various producers and firms. *Selection* takes place on the market performed by different user groups (consumers) showing their preferences choosing among the marketed moped types – thus establishing the eight stable moped families. The selection process is also influenced by the regulations and requirements put on the moped by the government during different phases of our history – i.e. the moped laws of 1952 and 1961. *Retention* is shown on the level of technical design in the stable moped families and on an institutional level by the persistence of the rule from 1952 that the mopeds had to have pedals like a bicycle – a rule that persisted a long time after the moment when technical design had made the pedals obsolete.”

Inspiration from evolutionary theory (and Biology) can also be applied to the sorting out of various technical artefacts by constructing what is called a “Phylogenetical Tree” by biologists and paleontologists. It is used when investigating family relations between different species. This method of sorting species or artefacts into family groups has been used by W. Bernard Carlson when analysing Edison’s sketches on the telephone. He and his associates defined every sketch as a “fossil”. By sorting the chronologically and looking for family resemblance they found both mechanical similarities and family ties on a more structural level:

“...devices that shared an underlying principle of operation, a common mental mode or meme. As we found telephones sharing an underlying principle of operation, we began to group them chronologically in horizontal rows on a map, letting them constitute different lines of [Edison’s] research”

Today’s evolutionary approaches to industrial and technical change may be looked upon as a third wave of evolutionary theory within social science. The first wave of *biologism*, emerged in late 19th century and disappeared around the period of WW1. That wave was by many labelled Social Darwinism although its *primus motor* was the sociologically interested engineer Herbert Spencer rather than Charles Darwin and had a strong influence on engineering communities and theories on industrial organization. This was also a period when evolutionary economics was founded by economists like

---


Alfred Marshall (industrial dynamics), Joseph Schumpeter (entrepreneurial/innovation theory) and Thorstein Veblen (institutional theory). Of these Marshall struggled with a strong biological influence.

The shortcomings of the early and narrow biological analogies in social science – before the evolutionary theory had matured and settled also among biologists – may have contributed to the difficulties for evolutionary perspectives to challenge the dominating equilibrium paradigm in economic theory in general but also in other social sciences. Although there are exceptions, it may be argued that the evolutionary approaches introduced by Schumpeter before WW1 as a foundation for innovation/entrepreneurship theory (the second wave) did not take off in theoretical development until well after WW2. And, as mentioned, this take off came with Nelson & Winter who showed the usefulness of general evolutionary thinking and analogies deliberated from 19th century biological overtones.

There was also – around WW2 – an important Swedish track of institutional and structural analyses with a clear evolutionary perspective; Erik Dahmén, among others, may be argued to belong to that tradition with his *structural theory of industrial and technical transformation* based on the *development block* concept, (Laestadius in this volume).

Today’s research on industrial and technical transformation (development, innovation, entrepreneurship and management – also organizational theory,) is thus based on a significant foundation of theories, partly conflicting, but to a large extent *related to an evolutionary perspective which implicitly or explicitly challenge the equilibrium dominance in standard economics and social theory*. We argue that researchers – and policy makers – with ambition to understand processes of industrial and technical transformation – and how to manage them – also must develop an understanding on the foundation of these concepts and theories.

Not the least is this important in a period like this when mankind is facing challenges to fundamentally transform the economy away from its addiction to carbon and towards sustainable life styles and means of production. Reducing GHG emissions with approx. 90% of the present will be necessary within a few decades. Old development paths have to be abandoned. New technological trajectories must be entered upon. New innovative models for business, for transport, for resource efficiency and for organizing the economy and society must be developed. This is probably the most far reaching process of industrial dynamics and transformation ever seen.
2: Eric Dahmén and Industrial Dynamics
Staffan Laestadius

Industrial and technical change takes place in a cognitive space and a field of forces where imbalances result in tensions that create incentives for change. Uneven development of an industrial transformation process may cause necessities as well as opportunities for actors in the system. This is the foundation for the theory of development blocks (DB:s) as developed by the Swedish economist Erik Dahmén. His analysis, inspired by Joseph Schumpeter, may be used to analyse the transformation of an economy – and its industrial and technical structure – on a meso level and without any assumptions of equilibrium which are common in orthodox economic theory.

The notion of development block catches the essential relations in an industrial/technological system but is not necessarily related to a certain industry (as defined in public statistics) or technology. The actors and forces of a DB may change over time as new technologies emerge and old ones fade away. This chapter not only introduces the DB concept and its theoretical embeddedness in detail to make it useful as a tool for analysis. It also illustrates how the DB concept can be used for managerial and policy purposes in the great transformations ahead for our society.

Introduction

This is a period when the world economy has to transform fundamentally in order to rapidly reduce the impact of modern industrial activities and life styles on the climate as well as to adapt to the climate change on the way. To a large extent that transformation has to take place within and between dominating firms and industries which during two hundred years of development have become highly dependent on fossil fuels. This transformation will – and must – impact on technologies and industrial capabilities and it will necessitate management activities as well as policy interventions.

To contribute to this change process, we need a theoretical understanding on how these systems fit together, how they develop, and on the mechanisms of change. There is the market of course, but industrial transformation is much more than just prices. This chapter introduces a classic Swedish approach to the study of industrial and technical transformation – the development block
analyses by Erik Dahmén. The basic argument – which will be revealed in the
text that follows – is that his approach still contributes with sharp and useful
tools for policy makers as well as for industrialists and
entrepreneurs/innovators for the understanding and change of our industry.

The idea behind this paper is that a qualified analysis cannot be performed by
just grasping a set of anonymous tools from the tool box. The advanced analyst
is also aware of the potential and shortcomings of the tools used as well as why
the tools were developed. Knowing this sharpens the analysis and reduces the
risk of walking in the wrong direction.

The chapter is organised as follows. Section two contains an analysis of an in
depth reading of Dahmén’s Ph.D. dissertation from 1950. Section three
focuses on the development block concept and how the Dahménian analysis
developed after his dissertation. In section four I discuss Dahmén’s impact
(and sometimes lack of impact) on the intensified research in industrial and
technical change which has taken place in the 1980s and the following
decades. Section five illustrates how the Dahménian concepts can be used to
analyse the industrial transformation and creation of new development blocks
that may take place during the decades ahead. This section thus illustrates the
most important managerial and policy implications of the Dahménian
approach.

**Erik Dahmén’s dissertation from 1950**

Erik Dahmén’s Ph.D. thesis – *Svensk Industriell Företagarverksamhet*
(Swedish Industrial Entrepreneurial Activity) – from 1950 is probably the first
empirical study ever based on a Schumpeterian analytical framework. It is
fully related to the mechanisms behind economic – or rather industrial and
technical – transformation. In that process Dahmén also introduces a set of
concepts which still are highly relevant for research on industrial and
technical change. These concepts may be looked upon as important parts of
the theory of industrial dynamics; in short: the analytical tools that connect
evolutionary economic theory with industrial and technical analysis.

The dissertation is written in a period when the academic discourse among
economists basically followed two tracks: on the one hand the main stream of
analyses was since long focused on the characteristics of and conditions for
equilibrium, on the other there was a strong focus on highly aggregated
models that (against the background of the crises in the 1930s) focused on
economic cycles. Not the least the second discourse (what Dahmén labelled

---

14 There are some other texts on the Dahménian approach. See eg. Dahmén (1991); Lindgren (1996);
Pålsson-Syll (1997) and Karlsson (2007). The intention here is to have a stronger focus on the
industrial-technical dimensions than is the case with those texts.
konjunkturteori – cycle theory) dominated Swedish economic analysis: even if John Maynard Keynes (Keynes, 1936/70) is the internationally most well-known representative for this track, this line of thinking also had strong roots among Swedish economists and within the Swedish political system (Berman, 2006).15

Dahmén positions his study in relation to business cycle theory which, he argues, in spite of advanced econometric methods does not get close to the core issues of economic development. He notes that new products, new methods, new forms of organization, and new markets – i.e. those innovations Schumpeter (1934) lists in his Theory of Economic Development – can destroy the old ones and still not be detected in the aggregates which cycle theorists analyse. Behind a vague up- or downturn there can be a significant structural transformation the tendencies of which business cycle analysis, with its focus on aggregate units like savings, consumption, wage level etc., run the risk of neglecting: “in such an analysis it has instead been most convenient to handle the transformation as datum” i.e. as something given – and not to explain (Dahmén, 1950, p. 5).

It may be argued that also John Maynard Keynes was well aware of these limitations in his General Theory. His discussion of “effective demand” is related to a “given situation of technique, resources and factor costs” (Keynes, 1936/1970, p. 24) and he also, in another of his works, makes a reference to Schumpeter as regards development in the long run, i.e. when technologies and industries are transformed or changes through innovative investments (Keynes, 1930/1960, p. 95f).

Dahmén also positions his analysis in relation to research in economic history, which, on the one hand has had historical transformation processes as one of its primary research objects but on the other, from various reasons, had not developed theoretical tools suitable for understanding the transformation mechanisms

As a consequence, following Dahmén, a gap developed between the theoretically advanced economists, who did not grasp or were not interested in the mechanisms of transformation, on the one hand, and the economic historians, on the other, who were interested in the processes of transformation but had not developed methods to analyse the causalities in these processes. That is the gap Dahmén intends to fill through analysing factual historical processes based on explicit, theoretical problems, i.e. what he labels a causal analysis. Although this analysis can be performed independently from the study of business cycles, Dahmén argues that the

---

15 For the use of sources and references in this chapter, see appendix A.
deeper understanding of transformation processes can contribute to business cycle analysis (Dahmén, 1950, p. 6f).

Three thought traditions or discourses on economic development may be argued to constitute the foundation for Dahmén´s dissertation. The first has its origin in Thorstein Veblen, whom in a critical way analyses the lack of evolutionary perspectives in economic theory, in the essay “Why is economics not an evolutionary science?” (1898) and in his book The Theory of Business Enterprise (1904). While the first essay is more general in nature, Veblen is in his book from 1904 more focused on the fact that entrepreneurial and industrial activities in themselves are lacking in economic theory. That was in fact what became the core issue in Schumpeter's dissertation from (1911/1934).

The second intellectual track for the Dahménian analysis is found in the research by his Ph.D. supervisor Johan Åkerman (also influenced by Veblen). Dahmén refers primarily to the more superficial text Ekonomiskt framåtskridande och ekonomiska kriser (1931) and the more theoretically developed Ekonomisk teori 1 & 2 (1939 & 1944) where Åkerman develops the distinction between “alternative analysis” and “causal analysis” (in Åkerman, 1944). While the former relates to the selection/choice between alternatives which economic actors are assumed to do all the time in an ahistorical context, the second concept relates to a historical process where time is important and the set of alternatives changes over time.

The third pillar in Dahmén´s analysis is the works by Schumpeter. In fact, Åkerman had put Schumpeter on the reading list for Dahmén in his Ph.D. work. If Åkerman contributes with the general background for the analysis Dahmén admits that Schumpeter´s text from 1911 contributes with the specific ideas. Not the least, Dahmén argues, this is the case with the “business aspect” put forward in Schumpeter (1911) where he makes a clear distinction between the technological dimension and the business dimension thus opening for “industrial activity” and “new combinations of production factors”. He also mentions Schumpeter´s skepticism towards aggregated analyses (Dahmén, 1950, p. 8f).

Dahmén describes the Schumpeterian analysis as a three-step process: first the comparison of the stationary and circular flow, secondly the disturbances in the stationary flow due to the innovations introduced by entrepreneurs and thirdly the adaption processes taking place following from the new combinations introduced in the system and which may cause a crowding out of old technologies and businesses.

He discusses the fact that the Schumpeterian approach can be used in the analysis of business cycles – which is what Schumpeter does in his Business Cycles (Schumpeter, 1939) – but that the most important implication of
Schumpeter’s work is that he does not restrict himself to the aggregated analysis of most economic theorists but to the fundamental character of the transformation process: primarily, according to Dahmén, on the micro character of the process, i.e. in the transformation of companies, of technologies and industries (Dahmén, 1950, p. 10f).

Dahmén did not include Schumpeter’s *Capitalism, Socialism and Democracy* (Schumpeter, 1943) in his Ph.D. thesis work. As a consequence the Schumpeterian metamorphoses in the view on the character of the innovation process and the dynamics of capitalism prior to WW2 (often labelled “Mk1” and “Mk2”) is not mentioned and still less analysed by him.16

After two chapters of detailed descriptive texts of the economic development before World War 1 and the interwar period, Dahmén, in his fourth chapter, introduces some of his concepts and his research questions. A core issue is that his causal analysis is related to “a non-reversible process and not to the time-less cause-consequence relation” (Dahmén, 1950, p. 45). Basically this is what Åkerman wrote in his *Ekonomisk kausalitet* (Åkerman, 1936) and further developed in *Ekonomisk teori 2* (Åkerman, 1944).

The double character of the transformation process is discussed in detail: on the one hand the positive fact that innovations are introduced, on the other the negative fact that the old is destroyed: technologies become obsolete and firms are closed down. Basically this is what Schumpeter (1943, chpt VII) labels “creative destruction” but that book is not referred to by Dahmén in his analysis. Methodological innovations (i.e. what we nowadays normally label process innovations) are analysed in relation to the standard concepts in micro economic theory: the production function and the isoquant. The choice between factor inputs following variations in factor prices but within the framework of a known technology is interpreted as a movement along the isoquant. Innovations, i.e. new methods, must – although it can sometimes be difficult to make a clear distinction – following Dahmén, be looked upon as a shift of the isoquant, i.e. a movement towards origo in the normal presentation of the production function as illustrated in fig. 1 (Dahmén, 1950, p. 46 ff).

---

16 In short: the further we moved into the 20th century it became obvious that the capitalist economy was transforming away from an “entrepreneurial capitalism” (MK1) into what may be labelled “corporate capitalism”; (MK2) i.e. became dominated by monopoly like structures created through processes of “creative destruction” combined with mergers and acquisitions. The Schumpeterian (1943) problem thus became how these great corporations with their monopoly power and bureaucracies could uphold their entrepreneurial spirits.
Explanation: X1 and X2 represent factor inputs. The curves Q1...Q3 represent different quantities (Q1 < Q2 < Q3) of a product or service that can be reached with various combinations of X1 and X2.

Two forms of innovations/technical change are illustrated. The arrow along the isoquant represents a shift of technology (the factor mix) to achieve the same output (here Q1). The arrow towards origo from Q2 may be interpreted as moving the Q2 isoquant towards origo i.e. achieving the Q2 level of output with less input of both X1 and X2. Many innovations may include both these varieties.

Dahmén is clear that this discussion is far from unproblematic. The isoquant is (in standard theory) normally assumed to be known making it possible for firms to locate themselves along it depending on relative factor costs. In cases when the isoquant is not known this Dahménian distinction is not applicable. This problem – whether the isoquant is or can be known and/or can be smoothly followed – has been observed in many critical analyses of the production function (see eg. Rosenberg, 1976).

The analysis of the interwar period gave Dahmén a solid empirical ground for his distinction between advancing, stagnating and disappearing industries. For the advancing he makes a distinction between what he calls market suction and market expansion, where the former is connected to demand mechanisms outside the industry itself (i.e. what today often is labelled “demand pull”) and the latter is basically driven by internal mechanisms like methodological (process) innovations, within the industry, i.e. what we may
label “technology push”. As regards stagnating and disappearing industries Dahmén argues that what is interesting from his analytical point of view are those industrial transformations which take place due to new processes and products, not due to cyclical phenomena or simple “malinvestments” (Dahmén, 1950, p. 49ff). Following the clustering in time of several advancing industries/technologies is, following Dahmén, what drives business cycles. This is very close to the arguments of Schumpeter (1943) and is also one of Dahmén’s main arguments against the shortcomings of aggregate analyses in the understanding of these cycles.

According to Dahmén, the transformation analysis is an argument also for institutional analyses in relation to the process, i.e. studies of the economic and political structure and the social structure from a sociological perspective (like educational level, income and power relations etc.). It is also a motive for analyses of the industry structure (ibid. p. 52 ff). This is in fact an argument for what his supervisor Johan Åkerman label “structural analysis”. Here Dahmén also formulated his main research question: to analyse the interwar process of business formation, business development and closures in Sweden. His ambition is to put flesh on the bones for the theoretical reflections (assumptions) provided by Alfred Marshall and Joseph Schumpeter (ibid. p. 55f).

Although Dahmén already in chapter 4 (ibid. p. 52) identifies the tensions that may take place due to imbalances in the innovation process that aspect is not developed further in the theoretical chapters 1 and 4. These problems, and the concept development block (in this text frequently labelled DB) are introduced for the first time in chapter 5 more or less ad hoc in the discussion on the unbalanced nature of economic development.17 Based on the very detailed analysis performed by Åkerman on the first half century of Asea (Åkerman, 1933) Dahmén concludes that Asea solved its balance problems in promoting subsidiary companies to engage in electrification of Swedish industry thus creating an industry in need of large power systems: “first through completing a full electrical “development block” did they manage to successfully create an electric industry” (Dahmén, 1950, p. 66ff).

This is also the chapter where the incentives for the transformation of development blocks, what Dahmén label “structural tensions”, are analytically developed in the dissertation. Dahmén is clear that these tensions appear – and can be studied – on at least two analytical levels. The first level is related to company, industry or institutional level: structural tensions can in this perspective be analysed in terms of over production, malinvestments and cultural and market related inertia. This may create difficulties for innovative companies to establish themselves. This level also includes the bottle necks

17 Both varieties of spelling – “bloc” and “block” – are used in this chapter.
which may occur due to limits set by communication structure and/or limited local markets (ibid. p. 68ff).

Secondly Dahmén relates his structural tensions to *technological and organizational development*. Even if it, following Dahmén, can be difficult to make a clear distinction between the economic/institutional level and the technological/organizational one he is of the opinion that we – on this level – face the challenges of completing the development block through *bringing technology and organization into fitness* between each other. These later balancing problems, which Dahmén found important, has earned remarkably little attention within economic theory although they probably are as important than the former as regards incentives for economic change (ibid. p. 70ff See also chpt 3 in this volume).

This later variety of problems reveal themselves industrially in production chains which, as a consequence of technical change somewhere in the system become unbalanced. Dahmén illustrates with steel manufacturing, a technology which changed rapidly during the 19th century and thus created structural tensions on many levels within Swedish industrialization: “ingot steel processes could sometimes not easily be introduced because the blast furnace technology was not developed in parity to produce enough of cheap pig iron to feed the Bessemer converters or Martin-Owens” (ibid. p. 71).

In summary Dahmén is of the opinion that development blocks have a strong cumulative importance, not the least when companies learn to think in blocks themselves and also to finance whole blocks: this, he argues, created the prerequisites for a cumulative expansion of industry, often with high profits. In short, this may be interpreted as Dahmén’s more empirically based contribution to the analysis of economic development once introduced by Joseph Schumpeter (ibid., p. 74ff).

**The development of the development block theory**

The transformation analysis and structural tensions were core elements in the *thought style* of Erik Dahmén. These conceptual elements were also something he struggled with already in the late 1930s. In his licentiate dissertation the concepts seems to have been integrated to a coherent conceptual world which also included the concept “development block”.18 There were also the “positive” and “negative” aspects of the development process, i.e. the double edged phenomenon Schumpeter (1943) labelled “creative destruction”.

---

18 The details of this have not been checked in Dahmén (1942), however.
It is far from obvious how concepts develop in the history of ideas and when or by whom they finally are introduced. That is the case here. It cannot be excluded that the term “development block” has its origin in Johan Åkerman or in discussions between the supervisor and the Ph.D. candidate (cf. Pålsson-Syll, 1997).

The structural tensions, causing uneven development between and within sectors (and technologies) may be identified as a corner stone in the development block theory. In fact, this non-equilibrium process, related to the innovation sequences, is found also in Schumpeter (1939):

> “Industrial change is never harmonious advance with all elements of the system actually moving, or tending to move, in step. At any given time, some industries move on, other stay behind; and the discrepancies arising from this are an essential element in the situations that develop. Progress – in the industrial as well as in any other sector of social life – not only proceeds by jerks and rushes but also by one-sided rushes productive of consequences other than those which would ensure in the case of coordinated rushes.... We must recognize that evolution is lopsided, discontinuous, disharmonious by nature – that disharmony is the very modus operandi of the factors of progress ... Evolution is a disturbance of existing structures and more like a series of explosions than a gentle, though incessant, transformation.” (p 101f).

Dahmén early connected his development block analysis to the *ex ante* – *ex post* discourse which developed among Swedish economists during the 1940s. This distinction, or dichotomy, can be used in several ways: the most common is to – like Gunnar Myrdal – use it to describe the selection set economic actors are facing and the expectation based decisions they make ex ante (in advance) in the beginning of a period on the one hand and what occurs ex post (afterwards) when the real effects of all actors’ decisions have worked through the economy, and eventually oscillated to a new equilibrium. Dahmén’s approach was somewhat different, however. He connected his *ex ante* concept to the industrial entrepreneurs which identified advancing industrial development blocks and thus proactively through their investments contributed to their growth. The *ex post* concept he connected to the reactive actions based on actors reading of price signals which reflect emerging unbalances and, which in their searching for profitable niches, contribute to fill in empty holes (Dahmén, 1991, p. 140).

Erik Dahmén needed seven years to finish his dissertation after the licentiate. To some extent this was a consequence of his large empirical work, partly performed during his years at the Swedish Industrial Research Institute, IUI. It is, however, his conceptual development rather than his empirical work that has become used by later analysts of industrial and technical transformation. That conceptual world did also become more developed and more strictly formulated in the later, although short and few, papers which he published.
Two of these papers deserve some comments here (both available in Dahmén, 1991, p. 126-148).

The paper on “Schumpeterian Dynamics” (ibid. p. 126-135) is interesting from two aspects: the first because Dahmén synthesizes his inspiration from Schumpeter four decades earlier, and secondly due to its very strong focus on transformation and “dynamics” rather than growth:

“Schumpeterian dynamics is characterized primarily by its focus on economic transformation rather than on economic growth, defined as an increase in “national product”, “capital stock” and other related aggregates. It contrasts not only with Walrasian macroeconomic equilibrium theory but also with neoclassical and postkeynesian macroeconomic growth models. Though “dynamic” according to generally accepted terminology such models do not analyse underlying processes at the micro level and in markets but instead relations between a number of broad aggregates and the result of such processes ...” (p. 127)

Analysing the consequences and mechanisms of Scumpeterian innovations and their role in “creative destruction” Dahmén continues that

“... transformation thus includes both economic growth and decline but a conceptual distinction is instrumental. This is because transformation analyses focus on causal chains outside the scope of growth analyses, namely on disequilibria and chain effects created inter alia by entrepreneurial activities, market processes and competition as a dynamic force. The micro underpinnings of such analyses therefore differ from those of growth models where the main interest is in aggregates, such as investment and saving, productivity, income distribution ... Seen through Schumpeterian glasses, the micro units have no well-defined generalizable “propensities”, and they are not fully informed calculators reacting in a mechanical way to prices they cannot influence. Instead, firms continuously seek new information and often search for projects which, if carried out, exert transformation pressure on the markets.” (p. 128)

In the paper titled “Development Blocs in Industrial Economics” (ibid. p. 136-148) transformation appears as a historical process of sequential movements between possibilities (or opportunities) and necessities. It is defined as sequences of complementarities which through series of structural tensions, i.e. situations of non-equilibrium, may result in a continuous transformation rather than a final equilibrium. Dahmén also makes a distinction between the, according to him, static concept competitiveness and the, dynamic, concept transformation power. While the former relates to the conventional price competition which prevails in equilibrium theory the later relates to the transformation potential inherent in the development blocs. Basically these concepts, close to those of Schumpeter (1911/34), argue that the essential competition is not based on prices and costs but on innovation and transformation.
It may be argued that Dahmén is of the opinion that the elimination of structural tensions, i.e. the filling of the development blocks, may result in situations of balance and equilibrium. But nothing in the rest of his texts on this matter demands or even implies equilibrium. It is thus, in analogue with global weather, possible to imagine a system of continuous disturbances, which all the time tend to fill out existing disequilibria/tensions occurring in the system but which during these processes continuously recreates new imbalances on micro level. In fact, equilibrium theory – to which Schumpeter devoted his first chapter in his *Economic Development* (1911/1934) – plays a very limited role in Dahmén’s analyses.

In addition, it may be argued that the industrialists and entrepreneurs, which are in focus in the Dahménian analysis, are not restricted to the restrictive assumptions in the core theory of economics. They may make wrong investments, they may have tunnel vision, and they may be wrong in their market forecasts and selection of technologies. These very human and non-perfect actions contribute to the structural tensions and create incentives for other actors in the economy.

Already in his licentiate Dahmén criticizes Schumpeter that he – in spite of his evolutionary approach – still is too stuck in his equilibrium thinking to draw the full consequences of his view on the essential nature of industrial activity in the capitalist economy. Also here, in analysing the dynamics of industrialism, Dahmén is well in line with his supervisor (see eg. Åkerman, 1932, p. 49).

**The role of development blocks in modern theory on economic, industrial and technical change.**

New ideas and concepts often develop in several varieties and in many different contexts more or less independent from each other. That is also the case with the development block concept which has “family resemblance” with other analytic approaches. One is the metaphor model based on the concepts, salients, reverse salients and critical problems, introduced by Tom Hughes (1992; see also chpt 3 in this volume). In the Hughes model the salient – originally a fortification term indicating a (fortified) position ahead of the frontier – is a technological component which has developed faster than other components in the system and thus represents an advancement along the development frontier. The salient thus indicates a disequilibrium, a critical problem, providing opportunities for carriers of the technology/system ahead to exploit their position. The reverse salient in the Hughes terminology is a system component that may have faced strong resistance and is lagging behind the frontier. Analogously, the reverse salient creates incentives for the carriers of the lagging system to catch up.
These reactions among the carriers of technology on salients and reverse salients do not necessarily end up in equilibrium. Solving the bottle neck created by a reverse salient may well end up in an over shoot, a salient, thus contributing to sequential equilibrium which may follow a certain direction, path or trajectory.

The historian of technology Tom Hughes analyses technical systems rather than industrial ones but that distinction is not always easy to keep clear. Both Dahmén and Hughes illustrate their technical arguments with the well-known imbalances between the spinning and Weaving industry in the eve of the industrial revolution. And, as mentioned, Dahmén explicitly argues that the structural tensions on the technological-organizational level probably are the most important and also those most neglected by economists. Here Hughes and Dahmén are very close.

Analysing the similarities, and differences between Hughes and Dahmén takes us to at least two problems worth more in depth discussion. The first is the systems approach, the second relates to the actors – entrepreneurs/innovators/industrialists – which drive the system forwards.

First, and maybe most important, is the fact that the transformation process in both approaches is analysed – and can only be understood – at a systems level which is less aggregated than the whole economy. Such an approach is natural for historians of technology as Tom Hughes but, as revealed in section two above – far from obvious for economists. We are not here talking on “micro level” which in orthodox economic theory relates to a level when the unit of analysis is “firm” or “household”, but a level between the standard aggregates in economic theory.

In his analyses Hughes works with a systems concept – “technological systems” – incorporating the relevant system components which cooperate to the functioning of the system. In short Hughes argues that his technological systems are socially constructed artefacts and that “inventors, industrial scientists, engineers, financiers, and workers are components of but not artefacts in the system” (Hughes, 1997).

In the early 1990s innovation researchers inspired by Dahmén (and independent from, Tom Hughes) developed a similar systems approach with elements from Dahmén’s structural analysis and with a clear technological profile. In the paper “On the nature, function and composition of technological systems” Bo Carlsson and Richard Stankiewicz (1991), against a background of emerging neo Schumpeterian innovation research, make an attempt to identify technological system as “a network of agents interacting in a specific economic/industrial area under a particular institutional infrastructure or set of infrastructures and involved in the generation, diffusion, and utilization of technology. Technological systems are defined in
terms of knowledge/competence flows rather than flows of ordinary goods and services. In the presence of an entrepreneur and sufficient critical mass, such networks may be transformed into development blocks, i.e. synergistic clusters of firms and technologies within an industry or a group of industries.” (See chpt 4 in this volume).

The technological systems concept was further developed almost a decade later by Bo Carlsson (Carlsson, 2000). The conceptual development must be understood against the conceptual congestion which has emerged within what may be labelled the innovation discourse: national innovation systems, regional innovation systems, sectoral innovation systems, competence blocks, clusters, industrial district and various network approaches all claim to have relevance for the understanding of this field of research. As mentioned, these systems approaches are further discussed in chapter 4 of this volume. Here I restrict myself to the general reflection that it is far from obvious that these innovation systems approaches give a fair understanding of Dahmén. As mentioned in section 2 above, Dahmén already in his dissertation Identified two analytical dimensions in his development block concept: one more industrialist related and one more technological/organizational. The technological dimension is more clear and important in Dahmén’s own texts than is the case in much of the innovation systems approaches.

Let us so turn to the second dimension of similarities and differences between Dahmén and Hughes: the role of the industrialist/entrepreneur/innovator. For Hughes, system building is the important process, and the innovators play the main roles. Although Hughes does not use the term himself, we can conclude that Hughes’ innovator heroes are entrepreneurial, they innovate and their systems also develop into profitable companies/industries (cf. Hughes, 1989). Hughes’s actors are labelled and identified as system builders (cf. chpt 3 in this volume).

Here we can identify a difference as regards the unit of analysis between Hughes and Dahmén. Although the industrialists are present in the analysis – not only in the title of his magnum opus – the analytical focus is on the structure and system rather than on the role of the individual in the creation of the system. Here, and somewhat paradoxically, it may be argued that Hughes is more close to the young Schumpeter (1911) than what Dahmén is. But this difference should not be overestimated: while Schumpeter’s heroes are more close to launching the innovations on the market, Hughes’s are more close to developing them technologically.

Industrial analyses in our time have a strong spatial or territorial focus. Not the least is that reflected in the innovation systems traditions discussed in chapter 4 in this volume.
It is maybe too easy to connect Dahmén’s works to some of these tracks. Not the least the Scandinavian innovation system researchers have normally read Dahmén or at least heard about his analytical approach. There are, however, strong reasons to stay clear in the analytical comparisons. Dahmén’s development blocks have no explicit spatial/territorial connection. The Development block is related to structural tensions which to a large extent – although embedded in socio-cultural mechanisms – basically are of technical nature: a bottle neck as regards process technology may well be solved with the help of imported artefacts or solutions from China.

This a-regional understanding of the original concept “development block” becomes clear in chapter 16 in the dissertation where Dahmén explicitly discusses industrial companies in relation to the small villages where they have emerged (a very Swedish “bruks” history). In this chapter he introduces the concept “economic-geographical development block” – a variant on the theme but here, when explicitly pointing to the spatial dimension, he uses another phrase (Dahmén, 1950, p.375).

This is not unimportant. Development blocks and technological systems, whether in the form of Dahmén, Hughes or Carlsson are not primarily geographical concepts. They are defined on a technological/cognitive space with now necessary spatial topological anchoring. The spatial dimension becomes an empirical question and something which – if necessary – can be handled in the policy analysis and as a challenge for management.

**Development blocks, industrial transformation and climate change – on the managerial and policy implications of the Dahménian approach in our time.**

Although man made climate change has been analysed during more than a century the awareness of the dramatic implications for our life styles and our industry and technology is more recent (and still difficult for many of us to accept) After two centuries of growing CO₂ emissions in the magnitude of 3%/yr., mankind has to reduce these emissions with approx. 5-6%/yr. in the decades to come to avoid a disastrous increase of global temperature in the magnitude of 2°C above late 19th century level (see Laestadius, 2013 & 2015, for further references).

When dealing – on a more concrete policy level – with this coming transformation the Dahménian tool box is highly relevant; and not only for understanding of the industrial/technical connections in society but also to change and transform them. First of all: GHG emissions mainly have their origin in a set of energy transforming development blocks which may be the
targets for “bottom up policies” with a strong focus on new technology and systems solutions rather than restricting policies on the macroeconomic level (i.e. general taxation and pricing systems).

Secondly: the DB approach facilitates the identification of core relations/connections – economic, technological, cognitive – across the borders of traditional sectors and industries thus revealing the important parameters to focus on. Development block borders do not necessarily follow the borders of industry classification: the DB for pulp and paper may include related segments from many different industries and technologies. And, as mentioned above, DB borders may change when old technologies are substituted by new ones.

Thirdly: the DB approach is fundamentally related to the dynamics within the blocks: the tensions, i.e. the necessities and the opportunities created during the non-balanced transformation. This is probably a core issue in the years to come: a giant transformation of industrial systems and technologies cannot be a process in balance. Neither was this the case in our industrial history of the past nor can this be the case in the future. Transformation policy must not be based on a smooth and balanced change. Entrepreneurial/innovative activities will identify existing tensions as well as create new ones. This was Schumpeter´s idea with creative destruction and so is also the case with the approach by Dahmén.

Fourthly: the DB approach has no necessary relation to “green growth”, a concept which has become popular in recent years (OECD 2011 & 2015). The transformation of blocks can take place irrespective of whether the economy as a whole grows or not. At best “green growth” is an outcome of the sum of block transformations. And at best there are – on aggregate economic level – self-propelling growth mechanisms of handling the necessities and exploiting the opportunities. But there is no guarantee that the net effect of the transformation of the development blocks can be accounted for as green growth (Laestadius 2013 & 2015). In fact, as mentioned in section 2 above, the empty space left due to the neglect by macro oriented economists and economic historians, who did not deliver explanations, on the transformation processes occurring beneath the cyclical processes was a core issue for Dahmén in his early writings already. Also in his later writings (cf. Dahmén 1991 and section 3 above) he was clear that transformation was the core issue – not economic growth. And this is still the case.

This argument can be illustrated from three sectors we may identify as development blocks (or even systems in the meaning of Hughes): energy, transport and steel.19 We may also face necessities and opportunities related

---

19 Biomass is analysed separately in chapter 12 in this volume. The transformation of biomass may well have connections to the transformation we discuss here.
to these sectors which may turn them into new development blocks. In fact, I argue, this is not only something we may study (ex post) as academic analysts. We may also (ex ante) act as managers or policy makers and try to create or transform these blocks!

As regards energy the systems/bloc character is obvious not the least because significant parts (electricity) are connected via the grid. The grid is always in balance (frequency stability) due to the character of electricity production. Industrial activity and household life styles are based on the continuous availability to electricity historically provided by large scale power stations (hydro, nuclear, fossil fuels and biomass/waste). Climate change now forces us to close down all fossil fuels (of which Sweden already has very few) while biomass partly will be allocated to other uses; and this parallel to a situation when there is strong resistance to nuclear power. One solution to this problem in addition to more energy efficiency is a large expansion of sustainable forms of energy like wind power and photovoltaics. But the sun does not always shine and the wind does not always blow. So there is need for more.

We are thus facing strong tensions in this development block. The immediately available technologies of sustainable energy production do not fit the existing structure of demand. This creates necessities as well as opportunities for energy providers as well as energy users. We cannot in advance know in detail how this block will transform although we can foresee very high prices for the most extreme demand for stability among a few users, new technical solutions for adaption to less stability (in grid frequency) among many users to avoid price peaks and new creative solutions to transform the base load system. New technologies for balancing and storing energy, not the least in the combination of heat pumps, district heating, biogas and hydrogen may develop.

The transport block, which may be defined wide or narrow, shows similar characteristics. A widely defined block for personal transport will of course include all subsystems related to the more “narrow” automotive block (which in itself is far from small as 82% of all European personal transport is carried out with cars) plus other systems related to personal transport. Climate change mitigation necessitates that all fossil fuels in car transport are eliminated. The only long term sustainable solution for fuelling the car is electricity, in some cases in combination with biofuels. But that transformation must probably be combined with a sharp reduction of car transport, growth of other sustainable means for transport like buses and light distribution trucks, overall reduction of personal transport, new models for ownership and organization of transport related services etc.

Also this transformation will create tensions with necessities and opportunities for various actors in the automotive block, which constitutes a large part of the industrial sector in Sweden. In addition, these tensions will
also be present for all those households (not included in Dahmén’s DB concept) which objectively or subjectively base their life styles on cars using fossil fuels. And these tensions are not far in the distant: if the car fleet will be fossil free in 2030 – a declared goal for Swedish policy and an important step to reach CO\textsubscript{2} targets as formulated in the COP21 agreement in Paris 2015 – approximately 300 000 cars from the present fleet have to be scrapped yearly and substituted by other transport solutions in combination with an immediate stop of new registrations of fossil fuelled cars.\textsuperscript{20} Also aviation faces a strong necessity to transform. Swedes fly more than average in Europe and the growth of air travel is larger than European average. And aviation is a significant emitter of CO\textsubscript{2} which has to be eliminated.

\textit{The steel block} is, since the eve of Swedish industrialization one of the backbones of the Swedish economy. Sweden has a higher production of iron ore and steel per capita than most other countries in the world. We also have a larger export specialization in iron and steel than most other countries; still more remarkable compared to other industrialized countries. Swedish iron ore and steel is in addition of higher quality than the global average: in short Swedish producers get more money per ton sold than most other producers.

The steel block is also one of the most GHG-emitting sources in Sweden thus making it to one of the core problems in the Swedish transformation away from dependence of fossil fuels. The reason for this is that carbon is used for the chemical process in the reduction of ore based iron, not only for “energy proper”. Scrap steel, however, can be melted and transformed with electricity which can be based on renewable resources. In short: large scale blast-furnace based iron production is not compatible with climate change mitigation. The problem is that Swedish steel (and iron ore) are the probably least polluting process in the world, per ton delivered and still more per function performed. Closing down Swedish steel will thus, in a world that still uses steel, only contribute to a leakage effect and other more dirty producers to invade the open space (Kander, et.al., 2015).

Of interest now, from a Dahménian perspective, is whether these tensions in form of necessities and opportunities can contribute to new development blocks. Not the least is this a challenge for industrial policy makers and industrialists to proactively develop new paths for a post carbon Swedish industry. Let us illustrate how this may take place in a new development block around \textit{hydrogen}.

Sweden can probably invest in renewable electricity – primarily wind power and photovoltaics – for a production of additional 100-150 TWh/year. When the wind is blowing and the sun is shining the excess capacity can be stored in

\textsuperscript{20} Many of today’s cars can drive on biobased substitutes for gasoline and diesel. But there will never be biofuels enough to feed a transport system with the structure and growth of today.
many ways, one of which is hydrogen. This hydrogen can be used in many ways: one is top/reserve capacity for energy shortages, another is for parts of the transport system. A third, and potentially significant use is for the reduction of iron ore to steel.

As regards reserve capacity in the grid hydrogen will probably never be a large scale battery, but it may be developed for local systems (maybe complementary to biogas turbines) like hospitals etc. This is an opportunity which still has to be developed.

As regards transport, fuel cell technology is already a working technology for cars although presently more expensive than battery based electricity. It is, however, an alternative way to store energy for electric cars. Also this is an opportunity that may develop from large scale hydrogen storing in an unbalanced system with cheap renewable electricity. In addition, we do not know the potential for hydrogen in shipping. What we know is that electricity based on present battery technology is no future for ocean shipping.

So, back to the steel issue: a hydrogen based reduction of iron ore to steel is probably the only realistic sustainable alternative to carbon. In fact, the process will emit steam water rather than CO$_2$. It is also innovative: in reality it means allocating resources on a technology of tomorrow rather than spending money on Carbon Capture and Storage (CCS) to preserve a technology from yesterday.

What we see here is the potential creation of new interdependencies between changing technologies and transforming industries, new tensions in the form of necessities and opportunities. We can, *ex ante*, imagine that we are contributing to the formation of new development blocks and that the old ones may change in character. At this stage we cannot imagine the details, but as we know that climate change mitigation and adoption is necessary (there is no Business As Usual as climate change will transform the conditions for human activities) this kind of creative analysis, entrepreneurship and management, based on Schumpeterian and Dahménian concepts and theory may be a good starting point for action: i.e. on how to destroy the old carbon based structures and how to create new development blocks for the future.

We do not have to know the details that may emerge. As regards the hydrogen – steel block we are just in the beginning of an unknown process. Enough is to grasp the core imbalances that have to be addressed. The entrepreneurs, researchers, innovators and policy makers of the future will create new – hopefully sustainable – tensions and new business necessities and opportunities along the way.

Neither do we know whether this giant transformation will create aggregate economic growth. Following Erik Dahmén we focus on the transformation of
industry and technology which is the foundation of all economic change. Whether this ends up in growth is another question (Laestadius, 2013 & 2015).

**On the use of sources and references in this chapter**

In his dissertation Dahmén makes frequent references to his intellectual sources. These references are often mentioned in this chapter when the origin in and inspiration for Dahmén’s conceptual development is described and analysed. The sources/references mentioned in this text are, however, sometimes those – younger – editions which have been used in when locating Dahmén in his intellectual context and when checking his understanding of his *inspirateurs*. The exact editions used by Dahmén is available in his reference list. An illustration: Dahmén refers to Schumpeter (1911) which is the original German Jena edition. In this chapter I have used the updated and translated Harward edition from 1934 and reprinted 1962 (Schumpeter, 1934/1962).

References made by Dahmén in his text are not referred to as sources here unless they are cross checked for the purpose of this chapter.

**References**

Berman, Sheri, 2006, The Primacy of Politics: social democracy and the making of Europe’s twentieth century


Dahmén, Erik, 1950, *Svensk industriell företagarverksamhet*, Band 1 & 2, Stockholm: IUI. (For an English translation of volume 1: see Dahmén, Erik, 1970,


Laestadius, Staffan, 2015, “Transitions paths – assessing conditions and alternatives” in Fagerberg, Jan; Laestadius, Staffan & Martin, Ben., 2015


3: Systems thinking in Industrial dynamics
Pär Blomkvist & Petter Johansson

Systems thinking is an integral part of the fundamentals of industrial dynamics. In this chapter, we explore and discuss systems thinking, what it is and how it can be used to analyse technological, industrial and societal change. The relevance of systems thinking have increased during the last decades due to challenges such as globalisation and threat of climate change. These challenges demand holistic approaches that encompass several different interconnected areas, including government policies, technological development, user practices, etc. In this chapter, we present such holistic systems approaches and provide concrete examples on how to apply them.

Introduction

Systems thinking is a way to understand interconnected parts that form a complex whole. It provides a holistic perspective – including both social and technical parts – that is central to researchers within industrial dynamics and many other fields.

In this chapter we delimit our focus to how systems thinking and related theories has evolved over time and been actively implemented in social sciences generally and industrial dynamics specifically. After that we introduce and discuss two systems thinking related frameworks frequently used in industrial dynamics studies: Large Technical Systems (LTS) and the Multi-Level Perspective (MLP). We also present a framework where we show how LTS and MLP can be fruitfully combined.

The LTS perspective is typically used to study large physically connected infrastructures, characterized in the LTS framework as large socio-technical systems. The MLP perspective is used to study societal and technological transitions and are popular in studies on socio-technical change towards sustainability. In MLP studies the systems under investigation are not necessarily connected by means of an infrastructural grid.

Systems thinking has a long historical background, going all the way back to the philosophers of ancient Greece. In the following we present a brief and condensed overview of systems terminology and the historical development of
systems thinking, leading up to current frameworks used in industrial dynamics research today.

**Systems thinking**

The English word “system” comes from the Greek word *systema*, meaning “whole compounded of several parts or members” (according to Merriam-Webster, 2015). What does a *whole* mean in this sentence? Aristotle famously expressed that ‘the whole is greater than the sum of its parts’, meaning that a *system whole* exhibit qualities that each part of the system by themselves do not exhibit. Today we refer to this phenomenon as *emergence*, also popularly expressed as “one plus one equals three”. The more complex a system is the more difficult it is to intuitively predict emergence from the system.

To understand what a system represents it can be fruitful to start with a simple system that is characterised by few components with few interactions. These are systems that are easy to understand and that we can recognize from our everyday life, e.g. our temperature regulating thermostat at home or the cruise control in our car.

![Figure 1: A model of a simple system](image)

Figure 1 shows a simple self-regulating system with the transformation process of *inputs* to *outputs* regulated by a *feedback loop* and with a permeable *boundary* towards the *environment*. If Figure 1 was to describe the cruise control of a car then the *output* would be the car speed, which depends on the *input* in form of the intended speed set by the driver and the actual speed that is supplied to the *transformation process* through the *feedback loop*. 
It is easy to see how designed self-regulating systems, such as the cruise control, have a very practical hands-on use. The mathematician Norbert Wiener (1948) used the term cybernetics to describe these types of self-regulating systems characterised by a (semi-) closed feedback signalling loop, i.e. man-made physical systems designed to fill a specific purpose. Though simple man-made systems are quite practical, they fall short when it comes to describing more complex systems that we encounter in our everyday life.

A complex system is characterized by a high number of components & subsystems and high level of interactions between these parts. The simple system in Figure 1 does not constitute a complex whole in itself, as we have defined it, so as we broaden our scope to the car including the driver in the traffic system the complexity increases.

A complex system has equifinality, which means the ability to achieve its goal in different ways. Sometimes complex systems are described as being on the edge of chaos because the interactions of simple sub-systems can result in unpredictable and seemingly random outputs. A complex system is also characterized by hierarchy – as seen illustrated in Figure 2 – and showing degrees of robustness and ability to adapt to its environment.

Systems thinking spread in many different research areas during the first half of the 20th century. The biologist Ludwig von Bertalanffy, together with other researchers such as the economist Kenneth Boulding, formulated the General Systems Theory (GST) in an attempt to synthesize the view on systems from different research areas (Von Bertalanffy 1968).
A system is characterized by the **interactions of its components** and the **nonlinearity** of those interactions

- Ludwig von Bertalanffy

In the work on GST von Bertalanffy highlighted the **interactions between components** and the **nonlinearity** of these interactions in a system. He also argued that no system that is open to interactions with the surrounding environment, i.e. an open system, should be viewed in isolation (Von Bertalanffy 1968). GTS was partly a response aimed towards a **reductionist** approach that suggested that all parts of a system could be divided into its individual components and each be optimised separately.

**Operations research**

In the 1950s and 1960s systems thinking became increasingly popular. The promises held by systems thinking as a way to optimise operations and decision-making made a significant impact on management research.

The wave of enthusiasm for applied systems thinking came out of the success of Operations research efforts during the Second World War. System theories had been used to quantify, calculate and optimize war time operations such as ship fleet logistics, air bombing, radar defence and effective weapon systems (Hughes & Hughes 2000). Operations research, Cybernetics and adjoining fields promised ways to actually **control** large and complex technological systems – in fact the term Cybernetics is derived from the Greek word for the steersman on a ship (kybernētēs).

After the war operations research became widely used in industrial management. Operations researchers addressed issues such as supply-chain management, floor planning, staff scheduling, etc.

In the end the ambitions to control social and technical systems did not deliver as much as first promised, but Operations research still made a big impact on management research and its effects are visible on management studies to this day.

As an example, social systems scientists are still using a way to define system borders that was given by the operations researcher and system analyst West Churchman almost 50 years ago. As the borders of a system are not given by nature, it is the task of the analyst to define the delimitations of the system that the analyst sets out to investigate. Churchman's approach to define what was a part of the system and what was not was to ask the following two questions: (1) Does “it” matter concerning the systems possibility to fulfil its goals? (2) Can the system do anything about “it”? (Churchman 1968).
**System dynamics**

Another area, with its roots in applied systems thinking made during World War II, is System dynamics. Influenced by the advancements of Wiener and Cybernetics-research the pioneering computer engineer Jay Forrester developed the field System dynamics at MIT Sloan School of Management in the 1950's. System dynamics is an approach to understand the complex and non-linear aspects of systems through feedback loops between interacting components, and also flows, stocks and time-delays in the system. In 1970 Forrester created a system model to simulate the world as a socio-economic system (called *World1*). Club of Rome three years later published *Limits to growth* (Meadows et al, 1972) based on a later, more elaborated, version on Forrester's work (called *World3*).

The increasingly popular agent-based modelling (ABM) approach represents more recent advancements in modelling social systems. The ABM approach is not equation based or structured as the above mentioned system dynamics approach, but instead focuses on the behaviour of actors and their behavioural rules. A possibility with the ABM approach is to construct a system and then remove particular agents one by one to detect which effect they have on the system – something that would not be possible using system dynamics. The hopes that today is put on ABM approaches in different fields, e.g. in the energy field as seen in Bale et al. (2015), reminds of the promises that operations research held during the 1950s and 1960s.

**Socio-technical systems**

Another area where systems thinking have had a big impact is organizational development, with important advancements made by psychologists at the Tavistock Institute in the UK in 1950.

Researcher Eric Trist and his colleagues at the Tavistock Institute coined the term “sociotechnical system” to describe the organization of teams in the mining industry and the interactions between the miners and the technology they used in their work. By showing the importance of optimising both the social and technical aspects of work, i.e. not only focusing on how to optimise each working operation but also including the quality of the workers' worklife, they provided a more holistic work design alternative to the reductionist Taylorism-approach that was customary at the time.

Note that the Tavistock researchers used the term “sociotechnical systems” in connection to management and organizational development, which is different from the way the term is used in the LTS- and MLP approaches, which will be introduced below.
**System thinking in technological system studies**

In the end of the 1970s a new way of systems thinking emerged in the fields of social science and history of technology. The earlier applied systems thinking, with its aim of full control had lost some of its popularity.

This new interest in system thinking was not so much aimed at the practical control of systems, but on understanding and *analysing socio-technological systems in their wider political, industrial and societal context*. It was a reaction to the continuing growth of the interconnected technical systems that took place after the Second World War – and also a reaction to the previously discussed ambitions to control these systems using for example operational analysis and cybernetics. As mentioned, these ambitions to control had not delivered as much as they promised.

The new discussion on systems, was driven by researchers such as Bruno Latour, Michel Callon, Wiebe Bijker, Trevor Pinch, and Thomas P. Hughes (Bijker et al. 1987) as well as evolutionary economists Richard R. Nelson and Sidney G. Winter (Nelson & Winter 1982). The advancements made by these scholars have been an important base for both LTS and MLP studies.

**Large Technical Systems**

As mentioned systems thinking has been used in many disciplines. In this part we turn to a physical and concrete meaning of the word *system* and discuss large infrastructural systems – *infrasystems* (Kaijser 1994) – and their dynamics. We present the so-called Large Technical Systems (LTS) perspective that was introduced by Thomas P. Hughes in the beginning of the 1980s.

Hughes was a mechanical engineer turned historian. His view on technological development in relation to the growth of large infrasystems has had a profound influence in the fields of industrial dynamics and the history of technology.

**Thomas P. Hughes and the LTS-perspective**

Industrialized nations are dependent on infrasystems. We need them for food, transport, communication, power, heating, water, waste, etc. In fact, modern life as we know it would be impossible without them. The systems are built up of components such as technical artefacts (cars, electricity cables and roads); institutions (formal and informal) and actors (professional and commercial), connected in mutual interdependence. (Hughes 1987) It is important to notice that each system component in itself could be viewed as a sub-system in the
larger infrasystem. It is the task of the researcher to define the system borders when performing the investigation.

LTS: s is often centrally planned and managed leaving users and appropriators with little power. The systems are not easy to change and deeply embedded in society. Negative externalities, such as pollution and global warming effects, are difficult to handle. Some of our infrasystems are not well adapted to the goals of sustainable development.

It is important to understand that most LTS: s did not start as centrally planned systems. They were not large and connected at all in the beginning of their history. The historical origins of most, if not all, large technical systems include some form of cooperation among its users. In Sweden, for example, the development of the electrical and the telephone systems started as cooperatives or other forms of joint ventures of users building the networks from the bottom up (Kajser 1999; Nilsson 2011). Historian of technology Arne Kajser (2002) describes the same process in the history of the Dutch struggle against floods. He reveals the historical and institutional heritage created by the appropriators’ efforts to build water control systems and its effect on Dutch political culture. But in the end, the Dutch water system, like the Swedish telephone and electrical systems, was transformed into a centrally administrated and controlled, tightly integrated, national infrasystem. Management changed from bottom-up to top-down. The appropriators at the local level lost their influence on the journey towards a large technical system. This centralization of the systems was also due to market mechanisms such as economies of scale and a consolidation of the whole industry around the system, often creating a monopoly or oligopoly market.

In Thomas P. Hughes (1983) ground breaking study, *Networks of Power*, on the development of the electricity system in the US and its diffusion overseas to Great Britain (London) and Germany (Berlin), he established the so called Large Technical Systems (LTS) perspective (Bijker et al. 1987; Hughes 1983). Hughes stressed that a large technical system must be analysed as a seamless web of socio-political, economic, cultural, institutional and technical components and thus not only technical but socio-technical in its nature.

Hughes uses the concept of technological style, or rather sociotechnical style, when discussing the differences in system development between the countries under investigation. The basic technological components that were transferred from the US was not changed in any radical way, but the system design and the institutional framing (ownership, regulations, etc.) became quite different when the Edison system was built in London and Berlin.

An often cited definition on a large socio-technical system was presented by the sociologist Bernward Joerges: “Such a complex and heterogeneous system of physical structures and complex machineries 1) whose material
components are integrated or ‘linked’ over large areas and over time ... and, 2) supports or facilitates the operation of a large number of other technical system, the organizations thus linked together.”

The keywords are that the system is “large”, it is “complex” in the sense that many parts of different character, both technical and organizational, are included and that the parts are integrated or “linked”. In addition, it is required that the system in question is interconnected with other systems.

The basic idea of the LTS perspective is that technological development cannot be analysed properly by only investigating the individual component or the different sub-systems. Here the car (the artefact/sub-system) and the road transport infrasystem is a good example. When we buy a car or sit behind the wheel we become, as a researcher has put it, not only drivers, but we are “buying in to a complex road-, energy supply-, spare parts-, maintenance-, registration- and insurance-, monitoring- and legal system” (Summerton 1998, p. 22).

The evolution of a socio-technical system can be described with the classical S-curve – see Figure 3 – and divided into three phases: build up – expansion – maturity. At this last stage, the LTS becomes increasingly difficult to change. In the mature system, expansion slows down when markets are near saturation (Hughes 1983).

When a system grows and especially in the expansion phase it can give the impression of autonomous growth and system expansion seems inevitable – in Hughes words: the system gains momentum. From a market perspective system growth can be understood through its positive system externalities. An infrasystem gets increasingly valuable when more people get access to systems services and components – it is not very useful to be the only one having a telephone.

---

When witnessing for example the motor road systems global expansion or the explosive growth of the internet and mobile phone systems, it is easy to believe in technological determinism. But it would be a mistake for managers and policy makers to yield to this impression – in the words of Tom Hughes:

“... but such systems are not autonomous. Those who seek to control and direct them must acknowledge the fact that systems are evolving cultural artifacts rather than isolated technologies. As cultural artifacts, they reflect the past as well as the present. Attempting to reform technology without systematically taking into account the shaping context and the intricacies of internal dynamics may well be futile. If only the technical components of a system are changed, they may snap back into their earlier shape like charged particles in a strong magnetic field. The field also must be attended to; values may need to be changed, institutions reformed or legislation recast.” (Hughes 1983, p. 465)

After the growth phase, the mature system is deeply embedded in society and not easily changed or transformed. The large mass of an LTS – the system inertia – consist of the physical infrastructure (roads, grids, pipes, tracks, etc.), the organizations and institutions created for support, laws and regulations and not the least financial and industrial corporations and individuals (engineers, scientist, politicians), with huge vested interests in the survival and expansion of the system. A mature socio-technical system is often very conservative, its actors – called system builders by Hughes – unwilling to change. It is often noted that it was not the stage coach owners that invested in the railroad and not the railroad actors that promoted the automobile. This feature is often called systems culture: A conservative attitude towards change, among the actors promoting, building and living of the system.22

---

22 A concept used by Swedish historian of technology, Professor Arne Kaijser.
Thus, large infrasystems develop according to a certain path or trajectory. The inertia acquired by the growing systems creates a “path dependence”. The earlier choices by system actors, and also “learning by doing”, influence the possibilities given to future actors. Donald MacKenzie takes this a starting point when he writes: “I do not deny that paths exist in the sense of clear/sustainable patterns of technological change.” But it is not easy to see if the technological path depends on technology itself or if it resides in the views of the systems stakeholders: “They invest money, careers and credibility in being a part of “progress “and helps in this way to create the “progress” of a certain expected form. Because the predictions turn out to be correct, they predict with greater confidence.” MacKenzie argues that the seemingly natural character of technological development is because “… in a certain sense, a technological development path is a self-fulfilling prophecy... In retrospect, it looks like to the prevailing technology possesses the internal qualities that explain its success.” (MacKenzie 1993, p. 168).

An important concept-pair in Hughes large technical systems perspective are the notion of Tightly coupled and loosely coupled systems. In a tightly coupled system the connections between system components or sub-systems (technical and institutional) are very strong and especially designed for system purposes. The railroad and the electrical system are examples of this type. They are centrally built, planned and managed, access for other operators is very limited, the technical standardization of components is high and they have a top-down perspective as distribution systems of goods. The system builder or the system manager has one overarching goal: To avoid technical and institutional mismatch in the system – to align system components within the system. A tightly coupled system is highly sensitive to disturbance, if one component fails the whole system breaks down (low redundancy). Therefore “load factor management” is important to avoid stress during peak usage and distributing the utilization of the system to reach economic efficiency (economic mix). Simply put, a loosely coupled system exhibits a lower level of the characteristics mentioned, with sea transport as one example. The road system would be somewhere in between the two system types. Historically it has moved from a loosely coupled system towards a more tightly coupled system, especially since the introduction of the automobile, and even more so with the introduction of information and communication technologies (ICT), but it is still not a tightly coupled as the railroad or the electric system. Advancements in ICT has transformed systems management radically. In some systems, like electricity, ICT has opened up the tight physical coupling between components. Auto-mobility and Air Flight has on the other hand become more tightly coupled systems with the introduction of ICT.
Changing Large Socio-Technical Systems

Tom Hughes perspective tries to solve the problem of combining the apparent, and seemingly autonomous, system growth with the knowledge that systems evolve, mature and decline, that they are, as any man-made institution, subjected to contingent historical context. Socio-technical systems are changed or being replaced by other competing systems. Gas light gave way to electric light; road networks has changed dramatically to meet the needs of mass motorization and today we are living in time when our global telephone system is gradually replaced by mobile solutions.

In the LTS-perspective, system change is caused by forces, such as: new technology, new or changed market conditions, environmental concern, policy changes and institutional innovation or lack of primary products (like oil). Factors that Eric Dahmén would label “transformation pressure” creating “structural tension” (see Laestadius in this volume).

Figure 4: Illustration of a reverse salient and a salient

In Hughes' (1992) terminology, the change caused by transformation pressure, takes the form of a “salient” or a “reverse salient” (terms similar to Dahmén’s structural tension). These are metaphors borrowed from the military, denoting a bulge that is ahead or behind of a front moving forward, see Figure 4. The concepts are used to describe a system component or a sub-system being ahead or lagging, compared to other components and given the desired direction of the advance. Salients and reverse salients disturb the system and slows down growth. They give impulses to innovate, but only if they are recognized by key actors.

The fact, that system actors must recognize a reverse salient, is crucial when managing a system in periods of transformation pressure. But identifying a salient or a reverse salient can be difficult for actors deeply embedded in the system.
A good example of a reverse salient in the electric system, that just recently been identified by system actors, is the meter placed in our homes. Although it is said to be “smart” it was originally designed to only measure electric current as a cost for the consumer. If you produced your own electricity (by wind or sun) it was difficult to sell surplus energy to the grid. The meter would bill your input as a cost on your own tab. Note that the meter is a reverse salient from the point of view of actors that would like to change the electric system to allow for both download and upload. This was not included in the traditional business plans of the large incumbent electricity utilities. The meter can be identified as a reverse salient if you are a wind power actor wanting to connect to the network. The meter can also, from the point of view of other systems actors, be identified as a salient. Due to technical development in other areas (ICT/Internet) the meter has become a more advanced system component and thus pulling the system towards change. This salient feature is the possibility of the meter to provide information on energy consumption and user habits and its ability, when connected to the internet, to serve as the central hub for controlling energy usage – that is to make it possible for the consumer to time energy consumption to electricity prices and thus become a central part in the vision of the “energy smart home”. Another example of a present-day reverse salient is the battery capacity in electric cars, which affects driving range. The battery is a lagging component in the (electric) automobile system. But it is lagging only if the range of Otto-engine cars is defined as the norm. Only if you define auto mobility by the performance of the gasoline car the battery becomes the reverse salient. But if you take global warming into the equation, it is in fact the Otto-engine that becomes the lagging system component and the battery/electric motor turns into a salient.

In the following we will concentrate on discussing the lagging components, the reverse salients. From the arguments presented above its easy to see that a reverse salient is not a given or natural feature of a system, it is not a dormant failing component, like a ball bearing deteriorating and suddenly breaking. Thus, a reverse salient is not possible to detect, using only perspectives from inside the system and in line with the prevailing system logic. A reverse salient is, to a certain extent, a social construction as its definition depends on the point of view of different actors – a reverse salient is a fact in the eye of the beholder.

The system characteristics mentioned above has primarily been used in historical studies to analyse the growth of infra-systems and the stability of mature systems. In the following we present our own development of Hughes LTS-perspective focusing on system actors and change management in infrasystems.
In the eye of the beholder ...

In the management of infrasystems it is evident that history matters. And it matters in two distinct ways. Firstly, when establishing a new system, the old way of delivering the service, forces the newcomer to adjust. The old system is, in a very concrete way, *already there*, and its physical infrastructure as well as its organizational design influences any attempt to create a new or complementing system. This feature is often summarized in the words of Edison when he tried to exchange gas light with his new electric light systems. To optimize the design of the system and to meet customer expectations and established behaviour – to minimize customer switching costs – a new system builder had to “Dress electricity in a gas costume”.

Secondly, history and heritage is also evident within the existing system and influences the way its actors are able to meet the need for transformation. History creates both physical path dependence and path dependant behaviour on behalf of the actors in the old system. The physical system components are standardized and aligned with each other and system actors are used to deal with system operations and business models in a certain way. Thus, every established system has its own systems logic. When change is approaching, it is not even certain that the incumbent system actors are able to identify the way to handle the upcoming challenges. It is hard to think outside the systems logic with its established architecture of standardized components, stable hierarchies and successful business models and when trapped in a conservative “systems culture”.

The fact that “history matters” leads us to suggest that the existing heritage of the prevailing system must be taken into account when trying to adjust it to a new situation. This holds true whether or not the plan is to construct a novel system (or sub-system) or if the ambition is to expand the old system along the lines of its particular systems logic.

Building on Hughes, we would like to differentiate between two types of reverse salients, and thus the need to manage change, that must be handled by actors within the infrasystem. Inspired by (Christensen 1997) we call them *sustaining* respectively *disruptive* reverse salients. In the words of James Utterback the conservative reverse salient is *competence enhancing*, given the prevailing systems logic, and the disruptive reverses salient is (possibly) *competence destroying* (Utterback 1994).

Firstly, system builders/managers have to deal with the need handle tensions occurring in the daily operations of the system. System actors most often handle these issues by “incremental innovation” and they usually focus on defining the problems as a “load factor issue”. We call these day to day issues *sustaining reverse salients*, and they are in often identified as simple and obvious bottle necks in the management and maintenance of the system.
These bottle necks are easy to identify by the system actors. They have established methods to deal with them and solutions follows the general logic of the system. A sustaining reverse salient does not challenge the systems culture nor does it threaten the business model of the system.

Secondly, system builders also need to address issues emanating from more serious structural tensions. We label these issues **disruptive reverse salients**. By this term we mean larger and more diffuse issues compared to the simpler ones mentioned above. If these disruptive reverse salients stay unidentified and if they are not handled, they can in some cases threaten the very existence of the infrasystem: “When a reverse salient cannot be corrected within the context of an existing system, the problem becomes a radical one, the solution of which may bring a new and competing system.” (Hughes 1987, p.75).

This differentiation between simpler sustaining and more complex disruptive reverse salients was touched upon by Hughes in a similar way – but not made explicit. Hughes discussed two types of innovations – conservative and radical – that could be formulated as a solution to systems components lagging in one way or another. Conservative innovations, corresponding to our term sustaining reverse salients, are often produced internally within managing organizations and thus creating improvements to the existing LTS. But Hughes also highlighted that a radical innovation, corresponding to our usage of the term disruptive reverse salient, often is created by external actors and that the radical innovation may produce a competing system. The history of Large Sociotechnical Systems shows that “…independent inventor-entrepreneurs could be shown to specialize in identifying critical problems and related 'reverse salients' on broad technological fronts.” (Joerges in Mayntz and Hughes, 1988, p. 13-14).

Of course, the incumbent system actors can use another strategy to deal with a disruptive reverse salient, in our expanded definition of the term, and that is to view it as an “externality”. That is, to draw the border of the system in such a way that the issues related to more serious transformation pressure is placed outside their responsibilities. But what happens when system actors must handle issues that they earlier saw as problems outside the realm of the system? These are the issues facing for example transport systems based on fossil fuels. Carbon dioxide emissions has been transformed from an externality to become a disruptive reverse salient that must be managed by system actors.

Tom Hughes stress the importance of identifying, in our wording, disruptive reverse salients and trying to formulate a solution within the system logic: “Defining a reverse salient as one or more critical problems is in itself a major step toward a solution, or invention, for it is well known that the ability to
define an amorphous situation as a problem is often an anticipation of a solution” (Hughes 1992, p.100).

But who has the right to define a disruptive reverse salient and point to a solution? In our interpretation of this process we want to point to the privilege of problem formulation. Which actors gains the privilege to formulate the problem and the strategy to solve it?

One recent example is the transformation of the aviation system. Low cost airlines such as Ryan Air defined luxury, luggage, comfort, etc., as disruptive reverse salients and removed these components from the system. At the same time the incumbent actors viewed all of these components as necessary and well-functioning system parts, and were not willing to change their business model. In fact, the low cost airlines managed to capture the privilege of problem formulation and they redefined air travel from being associated with comfort and leisure, into a simple and utilitarian commodity, such as a bus ride.

The Multi-Level Perspective

Over the last years the work by Hughes has received increasing attention as the threat of global environmental change has put increased focus on research concerning societal and technological change. Hughes formulation of a concrete analytical tool based on systems thinking is part of the theoretical basis for the new research area called 'sustainability transitions'.

This research explicitly aims to analyse societal and technological change towards increased sustainability. The challenges of achieving technological and societal transitions at regional, national and international level are of such complex nature that they require comprehensive systems approaches. One of the currently most prominent systems based approaches in the 'sustainability transitions' area is the Multi-Level Perspective (MLP).

The MLP approach was developed by Geels (2002) and Rip & Kemp (1998), and has since been further refined to constitute an applicable cross-disciplinary governance-oriented tool for steering societal change towards increased sustainability (Geels 2011). MLP studies divides the studied area/sector into three system levels (niche, regime and landscape) to explain the dynamic change processes between the different levels causing technological and societal transitions. The approach is illustrated in Figure 5, and is described as following by Frank Geels (2014):

“In a nutshell, the core logic is that niche-innovations build up internal momentum (through learning processes, price/performance
improvements, and support from powerful groups); changes at the landscape level create pressures on the regime; and destabilization of the regime creates windows of opportunity for the diffusion of niche-innovations. The alignment of these processes enables the breakthrough of ‘green’ innovations in mainstream markets where they struggle with the existing regime on multiple dimensions (economic, technical, political, cultural, infrastructural).

As previously stated, the ideas of Hughes are very much present in the MLP framework. But there are some important differences to highlight. Both the regime-concept used in MLP studies and the LTS-concept represents socio-technical systems. But where LTS represents a (relatively) tightly coupled physically interconnected technical systems (with social components) the regime-concept constitutes a non-physical structure that accounts for the stability of socio-technical systems. A regime does not necessarily represent a infra-based system (as LTS does) but is a representation of interconnected political, economic, cultural, social, institutional and technological elements (Geels 2004). The MLP approach also has a wider range of applicability, (cf. Arvidsson in this volume). The socio-technical regime also accounts for a selection-retention mechanism for different innovations (see the introductory chapter in this book for more about variation-selection-retention mechanisms).
The other difference between LTS and MLP is the explicit division of the studied area into three system levels to analyse the dynamics. Coevolution of actors and institutions take place at different system levels and is influenced by variables that change over time (Lewin & Volberda 1999), and the MLP framework provides a structured approach to deal with these multi-level co-evolutionary processes.

In the following part we will discuss each level individually and then provide some examples of MLP analyses to describe the interactions between the relatively stable socio-technical regime and the “flurry of change activities” at niche level which takes place in the regime's environment represented by the landscape (Geels 2012).

**Niches**

As explained by Geels, innovations and novelties are often developed in niches, which are protected environments where the innovation can mature and be tested before being included in the system. The niche level is derived from the idea that radical novelties (innovations) are developed in niches that enable cross-disciplinary experimentation. The niches enable specified interactions between issues and actors. Well-developed niches are suggested to act as building blocks for change; they are central for regime shifts.

An example of a strategic niche is a military research project which is not exposed to the selection-retention characteristics of the market. In The Entrepreneurial state the author Mariana Mazzucato (2013) brings up a string of examples of important technological innovations in the iPhone – Apple’s flagship product – that have been developed in government-protected niches. These includes innovations such as the internet, the GPS, the touch-screen display and the voice-activated ‘Siri’. Another niche, in the form of a so called protected market space, is exemplified in the steam ship example below.

Niches are often subsidised to support non-profitable innovations with the expectation of future societal benefits (Schot & Geels 2008), as the examples from Mazzucato shows. However, niche innovation may be misaligned with existing technological infrastructure, regulations, user behaviour, etc. Factors in the surrounding system are crucial for an uptake of innovations, and large-scale changes (or “regime shifts”) (Schot & Geels 2008).

**Socio-technical landscape**

Geels (2002) argues that radical innovations developed in niches can make a break through when the stability of socio-technical regimes are confronted with problems and tensions and the links between the elements within the regime are “loosened up”. This can for example happen through external...
pressure from *landscape level*, i.e. the level that constitutes the environment of the socio-technical regime.

As explained by Schot and Geels (2007) problems as well as opportunities can emanate from factors stemming from the landscape level, which are not controlled by actors on the local (niche) level nor at regime level. Changes in the landscape can influence the system development in ways difficult to discover for the actors involved in its daily management (Blomkvist & Larsson 2013).

The sociotechnical landscape is equivalent to what other system theorists often refer to as the system's environment. In Hughes (1987) words the system environment is the societal context of the system. The sociotechnical landscape is according to Schot and Geels (2007) characterised by the set of rules that guide technical design, the rules that shape market development and rules for regulating these markets. The socio-technical landscape includes the institutional and market aspects required in order for lower levels to function.

**Socio-technical regime**

The term “technological regime” was first introduced by Nelson & Winter (1982) to represent a set of shared beliefs among engineers and to describe trajectories of innovative processes in industrial sectors.

Nelson & Winter explained the concept of technological regime using the airplane Douglas DC-3 as an example. When the DC-3 was introduced it had new features, such as all-metal skin, a powerful engine (a DC-3 could tow up to three transport gliders at once), a low wing and streamlined body, etc. These features influenced the whole industry as a basis of how an airplane should be designed. In other words, the technological regime of the DC-3 shaped the development trajectories and delimitations of future technological change in aeronautics (Nelson & Winter 1982).

The discussion concerning the social construction of technology (SCOT) during the 1980s, where the work by Hughes was central helped further develop the concept of technological regimes to include social actors such as researchers, policy makers, users and associations as well (Bijker 1995; Geels & Schot 2007). The term 'technological regime' was expanded and got the new epithet 'socio-technical regime', which strengthened the perception that regimes (technical or socio-technical) stabilized current development trends in several ways (Geels & Schot 2007).

According to Geels (2004) a socio-technical regime consists of three interlinked dimensions: (a) the network of actors and social groups, (b) formal, normative and cognitive rules and (c) physical and technical elements.
Examples of formal rules are building regulations and laws, examples of normative rules are norms of behaviour, and examples of cognitive rules are guiding principles, how a problem formulation is made, belief systems and heuristics (Verbong & Geels 2007).

A socio-technical regime affects development paths both for technicians and companies and can be compared with the concept innovation system – see chapter by Rickne & Laestadius in this book – which includes elements such as actors in the value chain, networks, institutions (which here means rules, norms, cognition) and technology (Jacobsson & Bergek 2011).

**Examples of socio-technical regime reconfigurations**

To illustrate the interaction between regime, niche and landscape level Frank Geels (2002) uses the example of steamships. Steamships were a radical innovation that made a breakthrough in a reconfiguration process of socio-technical regimes that happened due to landscape pressures.

Steamships began to appear in the early 19th century when sail ships dominated the overseas transport regime. Steamships had reduced cargo space compared to sail ships and were driven on expensive coal, but had the advantage of being able to provide line services with both a fixed departure and arrival time, which the wind reliant sail ships could not. Steamships were therefore mostly used for personal and mail transport in the beginning, as well as for high-value low-volume cargo, where speed and regularity were appreciated.

At the time the socio-technical regime for over-seas transport was dominated by sail ships: the ports were made for sail-ships, the ship-builders only knew how to make sail ships, there was no infrastructure for steamships to reload with coal, the insurance companies did not want to insure the uncertain steamships, etc. This created stability for the sail ships while hindering the breakthrough of steamships.

Also, at the time, the performance of steamships was not that good. The initial steamships had paddlewheels and were made of wood. They were therefore most suitable to traffic inland waterways. At open sea, the paddlewheel did not stay in touch with the water all the time and the heavy steam engine machinery made the wooden hull bend. Experiments with making iron hulls in some cases resulted in steamships turning upside down once outside of its dock.

In 1838 however, the British government subsidised steamships to improve the speed of communication on special routes important to the British Empire. These subsidies created a “protective environment” for developing oceanic steamships. The results were increased experimentations with iron
hulls, screw propellers to replace the paddlewheels, more effective engines, etc. Thus, accumulating innovations within a niche.

However, these early oceanic steamships were in the beginning a complement to sail ships. They were in fact often sail ships with a steam engine as an “auxiliary add-on”. Early steamship advancements therefore rather strengthened than reconfigured the sail ship technological regime, in a hybrid state.

In the 1850s and 60s two events on landscape level gave steamships an unforeseen advantage. At the time, most passengers preferred the quicker steamships compared to the time-uncertain sail ships. The emigrant-wave from Europe to America became the first major market segment dominated by steamships. This emigrant wave was driven on by famine and political revolution in Europe, as well as the Californian gold rush. Events that were clearly outside of what the actors in shipping and overseas transport could affect.

Another important landscape event was the opening of the Suez-canal in 1869. The canal was unsuitable for sail ships and therefore gave a great advantage to steamships on the route Europe to India and China.

By the late 19th century steamships finally overtook sail ships as the dominating alternative for overseas transportations. Note that the shift from sail ships to steamships took almost a full century and was a stepwise process of reconfiguration. Partly because the sail ships, as a defence mechanism, adopted several radical innovations such as iron hulls (to increase cargo space), on-board machinery (to reduce need of manpower), etc., which increased the performance capacity compared to steamships.

So, the success of steamships was not solely dependent on technological advancements of the steamships themselves, but dependent on a set of developments at multiple levels that would link up and allow for steamships to break out of its subsidised niche. In Geels own words: “Breakthroughs of innovations thus depend on processes on the level of regimes and landscapes, i.e. they are context-dependent. It is because of this aspect that the multi-level perspective is useful for analysing technological transitions.”

Another more modern example of how stable socio-technical regimes can change is seen in the Swedish residential heating sector and the shift from oil boilers to the heat pump technology.

In the early 1970’s oil boilers dominated the Swedish residential heating regime for small houses. The oil crises in 1973 and 1979, causing a quadrupled price on oil, provided strong incentives for the Swedish government to reduce the dependence on imported oil. The Swedish government launched a series
of subsidy programs to develop heating alternatives such as solar technology and heat pumps. The state-owned power utility Vattenfall was assigned to be an active part of the technological development. Vattenfall quickly turned their attention towards heat pumps instead of solar energy, as the heat pump fit better into Vattenfall’s electricity system. In a few years over a hundred companies joined the new and seemingly lucrative business of producing, selling and installing heat pumps to house owners. Many of these companies received vital support from Vattenfall. In this niche the heat pump technology was developed to match the Swedish climatological and housing conditions (Swedish houses are typically fitted with water-based heating systems). However, the existing heating regime was not aligned towards heat pumps. Problems occurred of both technical and social character, e.g. problems with house mould due to less natural ventilation without the warm oil boilers and lack of education of installers which caused substandard installations. As these problems were gradually mitigated and the performance of heat pumps increased the heating regime for small houses in Sweden stepwise reconfigured towards heat pumps instead during the 1990’s. Today there are more heat from heat pumps per capita produced in Sweden than any other country. (Blomkvist & Johansson, 2014)

**Multi-Level System Diagnosis: Combining LTS and MLP**

In this last section of our chapter we would like to introduce a framework used when investigating and analysing change processes in infrastructural systems. The framework focus on the system actors point of view and on finding reverse salients. It is primarily based on Hughes LTS-perspective and, in early stages, developed independently of the work of Frank Geels. The framework has since then been strongly affected and inspired by Geels and MLP.\(^{23}\)

The framework divides the LTS under investigation into three levels and the purpose is to locate reverse salients on these different levels. These reverse salient, we argue, causes misalignment between system components and slows down system expansion. Following from the discussion above, we argue that a functioning large socio-technological system needs alignment between different components within the system and alignment with other systems and institutions in society. The purpose of the diagnosis is thus to identify

misalignment between systems components or sub-systems. Each level has a couple of general research questions to guide the investigation:

- **Local alignment**: How do we organize and manage the LTS at the local level? Are there reverse salients causing misalignment on the local level that could cause the system to slow down its development? Is there misalignment between local system components and the next level of the LTS – the sociotechnical regime?
- **Sociotechnical regime alignment**: How do we organize and manage the LTS at the sociotechnical regime level? Are there reverse salients causing misalignment on the regime level that could cause the system to slow down its development?
- **Landscape alignment**: How do we align the whole LTS (local and regime level) with neighbouring institutions and systems in society? Are there reverse salients causing misalignment on the landscape level that could cause the system to slow down its development? Problems that threaten the development of a LTS can arise not only from within the system (local and regime level) but also externally. Conflicts can stem from collision with other societal institutions, rules, organizations, or systems. Hughes (1987; 1992) uses a similar distinction when discussing the delimitations of a large technical system, which he calls environment. Thus, our third analytical level will focus on friction between the two first levels in the LTS and its environment. But a note of caution is appropriate before moving on. The “landscape” of a LTS can be many things. Grand societal changes like industrialization, urbanization, and modernization are concepts that can be used to contextualize changes on the landscape level. However, in our third analytical level our purpose is not to address these types of general and abstract concepts. We want to keep “landscape” on a fairly concrete level. When talking about “landscape misalignment” we mean external factors causing direct friction when interacting with the LTS under investigation.

In the following we will give an example of how to use the diagnostic framework on a specific case: wind power establishment in the existing electrical system (in this example we do not differentiate between sustaining and disruptive reverse salients). We use local, sociotechnical regime and landscape alignment to analyse the ongoing wind power establishment in Sweden of today (based on Blomkvist & Sandberg 2011).
Wind Power and Local Alignment

Actors in the wind power industry experience problems in getting acceptance for wind power projects in the local community in rural areas. In many instances the problems originate from lack of experience and routines and a lack of incentives on the local level.

Firstly, one important reverse salient is that inhabitants in the rural areas has no direct experience in working together in large and capital intensive projects. Knowledge and economic capability is a scarce resource. There is no established arena where discussions can take place and the local community (i.e. the village) has no clear and accepted institutional framework to handle these complex questions and no legitimate procedure to arrive at a decision. In short: the village is not a well-defined actor.

Secondly, there are no clear economic incentives to support the project if you are not the actual landowner. The windmill affects the living environment for many, but just a few gain economically, i.e. positive and negative externalities are distributed unevenly. On big obstacle is the “Catchment area of wind power”. When building a wind power-plant you inevitably reduce the possibilities to build recreational facilities, houses or other windmills in the area. The plant creates “wind shadow” for other windmills and reduces other landowner’s ability to exploit their land. Actors in the industry try hard to innovate models for partnership, sharing of profits and cooperative ownership.

One way to solve these reverse salients and improve local alignment is to create an organization – a distinct actor – similar to the already existing road associations managing local roads. It is probably both economically and socially profitable to establish local institutions for cooperation in wind power projects. Local alignment appears when all sorts of economic and social externalities, revenues and costs, are shared in a fair way. Local alignment can also reduce “NIMBY-sentiments” in the community (Devine-Wright 2009).

Wind Power and Socio-Technical Regime Alignment

Local misalignment is not the only problem of wind power establishment. Wind power mills are small compared to the centrally placed distribution facilities in the existing electrical system and they do not always fit in the distributive network managed by the electricity utilities.

Firstly, there is number of technical reverse salients causing miss-alignment between the old system end the newcomer. The Swedish electricity system has been designed as a distribution system from the centre to the periphery. Wind powered electricity is small scale with decentralized production plants. A new
grid design needs to be implemented to allow for input of electricity into the grid from windmills (i.e. “smart grid”).

Secondly, there are economic reverse salients in the difficulties of billing, due to the contingent nature of the wind. A windmill owner can sometimes deliver electricity to the grid and needs to get paid for this contribution. At other times, when there is no wind, the owner must be able to use electricity from the grid and consequently pay for that service.

Thirdly, institutions, regulations, organizations and markets are designed exclusively for a distribution system with large central production plants. It is hard to meet rules on security and reliability for smaller producers. And furthermore, the dominating players, de facto an oligopoly, has no interest in letting new players enter the market.

Concerning the socio-technical reverse salients and miss-alignment our advice, derived from historical example, would again be to learn from the road sector. During most of its modern history, the local roads have been an important issue for the state and for the state road administration. Special technical personnel were given the task to monitor and advice local road associations. The state road administration, with its primary mandate to manage public roads, established regional offices where designated road engineers and legal experts handled questions of socio-technical alignment. We believe that the wind power sector would benefit from these types of arrangements.

**Wind Power and Landscape Alignment**

One surprising miss-alignment appeared when the communal income tax law from 1928 was rediscovered, and unexpectedly, became a reverse salient in the wind power system. It was originally aimed at farmers using surplus goods, e.g. milk, meat, and eggs, from their business to feed themselves and their families. The tax law postulated a stereotyped tribute for every farmer adding the value of the benefits in kind to the yearly income. By a decision by the central tax authority in 2008 the same rules were to be imposed on collective ownership of wind power. Members of cooperative societies owning shares in a windmill had to pay tribute if they were able to produce their own electricity at a lower cost than the market price at Nordpool spot market. The direct result was a 90% decrease in joint owned windmill projects.

It is of course possible to identify reverse salient causing landscape miss-alignment regarding laws and customs from other areas, i.e. public right of common (allemansrätten) animal and nature protection, cultural heritage, etc. There are also many organizations or groups of people outside the local or sociotechnical regime level that interpret wind power as a threat to their interest, i.e. bird and nature organizations, recreational and sport
organizations and groups setting out to protect the rural landscape. To avoid landscape miss-alignment it is important to investigate and diagnose the environment, in a wide sense, of the wind power project.

**Managerial and policy implications**

Systems thinking has a wide range of applicability and the importance of systems thinking seems to increase every year. Many of the ongoing challenges related to globalisation and climate change involves structural problems that demand holistic approaches where the wellness of “the whole” – and not only the individual parts – is in focus.

Systems thinking is thus important for both policy makers and managers. A problem that relates to complex interactions of different social and/or technical elements can never be fixed by focusing on only one of the elements: all parts of the system must be addressed in order to find a fulfilling solution. The usefulness of system thinking is that it is applicable to both small and large systems – it can be equally useful to use for the project manager at the small company as well as for the politician engaged in bilateral discussions on a nation’s energy supply.

The LTS and MLP concepts provides concrete tools for analysing changes and transitions in large socio-technical systems. By understanding the concept of dynamic stability of socio-technical systems the manager and politician better understand the challenges for engaging in transition activities. It also provides insights to the importance of multi-level engagement. It is not enough with a new energy-efficient technology developed in a niche if the stability of the regime blocks the technology from achieving widespread use – alignment is crucial.

**On methodological considerations and the use of sources in this chapter**

This chapter is mainly based on a literature review. But we have also included exemplifications from our own research.

The scientific foundation on general systems theory is presented in Anders Karlqvist i Teknik och Samhälle. En systemanalytisk introduktion, Tema T Rapport 4, 1983, Universitetet i Linköping. We have also used Hughes, Agatha C. and Thomas P. Hughes (eds.) *Systems, Experts and Computers. The System Approach in Management and Engineering, World War II and After* The MIT press, 2000 and the chapter by Arne Kaijser and Joar Tiberg in this


The MLP perspective has generated a lot of research in the last decade. We would like to point at one of the most influential articles that can be said to have initiated this field: Geels, F.W., 2002. “Technological transitions as evolutionary reconfiguration processes: a multi-level perspective and a case-study.” Research Policy.

Suggested further readings

- The Social Construction of Technological Systems: In this seminal book a discussion, together with several important insights and ideas about technology development, is captured. In this book the concepts of Large technical systems (LTS), Social Construction of Technology (SCOT) and the Actor-Network Theory (ANT) are presented (Bijker et al. 1987). Concepts that still are of relevance for both managers and researchers still.


References


Introduction

The aim of this chapter is to scrutinize the theoretical foundations of innovation systems, which is a conceptual framework for the analysis of innovation processes and innovation policy within a certain context (defined as a system). The chapter reviews the state of the art, i.e. our ambition is pedagogical and synthetic rather than to be fundamentally critical or to develop theory. In extension, the aim is to sharpen the intellectual and analytical tools of innovation systems, to make them still better and useful for the analyst to understand – and handle – the great industrial transformations we have reasons to expect for decades ahead.

Let us, for the sake of convenience, already at this stage provide a broad definition: an innovation system (IS) is, in this chapter, defined as an interrelated structure of institutional and actor based condensations in an economic space. These actors are engaged in generation, diffusion, and utilization of innovation, in an area which has specific industrial and innovative properties compared to its context with which it has exchange. Commonly, the components of an IS are defined as all the various types of actors within the system which are relevant for innovation processes to come
about, the networks and knowledge flows between them, the knowledge areas (i.e. technologies), and artefacts involved in the specific system, and the institutional set up guiding the behaviour of the components as well as of the system as such. The various innovation systems approaches – be they sectorally, technologically, nationally or regionally defined – thus focus on the fact that emergence, specialization and competitiveness of sectors/knowledge areas/countries/regions relate to the institutional conditions, the setup of actors and their specific, path-dependent, knowledge base and how this is shared.

The chapter is structured in the following way. We start with laying the foundation on which these system concepts rests: an evolutionary view of innovation (section 2). Indeed, we see modern innovation theory as a reaction to the often static views of economic theory with its focus on equilibrium rather than dynamic changes, selection mechanisms, learning processes and path dependencies. Next, in section 3, we return to the innovation system concept presenting its basic characteristics, paying particular attention to the importance of institutions, the problems of openness and intentions in systems, as well as the identification of systems borders.

Paving way for a deeper understanding of the policy uses of the IS concept, we dig, in section 4, into the various IS related concepts. That section is primarily of a typological character with the intention to give an overview of the conceptual jungle of partly overlapping concepts.

Section 6 focuses on two problems for which the IS approach has been criticized. First, one set of critique says that in spite of claiming to address evolutionary aspects of innovation there is a risk of rather static and structure based analyses. Several authors have addressed this critique by developing a functional approach to the analysis. Second, critiques assert that while the institutional set up is used to delimit the systems analysed, institutional changes and influences, as well as governance aspects, have more seldom been captured by IS analyses.

**An evolutionary economic view of innovation**

Our analysis focuses on the innovative properties of innovation systems (IS). We can in the definition of IS above find an implicit criticism of orthodox economic theory with its focus on equilibrium. Innovation theory, with a cultural heritage back to the Austrian economist Joseph Schumpeter, centres on the breaking up from equilibrium, on the creative destruction when new products and processes substitute for old ones, and on the transformation and growth taking place as a result (Schumpeter, 1934/62).
In the innovation literature, entrepreneurship and innovation are different, and complementary, sides of the same coin. The entrepreneur is the actor who introduces the innovations on the market; be they science based, technological, service related, organizational or market related innovations. In the literature, an entrepreneur without innovations is not perceived, even though in common jargon anyone stating a new venture can be called an entrepreneur. However, the strong interrelation between innovation and entrepreneur is an important part of the original theory pointing towards the action oriented element in the innovative economy, not only the exploration aspect (focusing on inventions and R&D).

Basically, innovation theory is related to what may be labelled *evolutionary economics*. The logic is simple: the essentials of the Schumpeterian message 100 years ago was that innovations continuously destroyed the equilibrium process and forced the actors of the economy to a behaviour which ended in industrial transformation and change. In evolutionary theory, equilibrium thus is a special case, an attractor on systems level for sure, but never realized as the actors in the system always strive to avoid “the circular flow of economic life” as Schumpeter formulated it (Schumpeter, 1934/62, chpt II). Competition in this world is not primarily a price competition between homogenous products among independent and fully informed actors, but a continuous struggle to use different solutions (e.g. advanced design) to create variety that attracts certain consumers. As an outcome this creates temporary monopolies in all areas, thus providing a short period of profitability before the competitors catch up with new solutions; more or less influenced by, or deviating from, the original ones (Schumpeter, 1942/2000, chpt. VII).

The attempts to combine the strengths of evolutionary theory with equilibrium theory have a long tradition among economists and in fact anticipate the contributions of Schumpeter. Most interesting, and still relevant, are probably the serious attempts made by Alfred Marshall in late 19th century already (Marshall, 1890/1990; Laestadius, 1992; 1999). He did not solve the problem of combining evolutionary theory with equilibrium; during the process of trying he, however, introduced the concept “externalities”, which became the cornerstone in his analysis of learning effects, knowledge spillovers and cultural influences, transmitted outside the market but over the borders between firms as “were they in the air”. For Marshall this was a core concept in his analysis of the economic and industrial dynamics in industrial districts. Although he did not use the concept “innovation”, what he analysed was in fact the – cultural and knowledge based – conditions for innovative processes.

That takes us back to the IS concept. Originating in the analysis of Marshall, there is an insight that externalities is one of the core concepts for understanding the learning and knowledge formation process characterizing dynamic systems; be they industrial districts or innovation systems of
“higher” order. On all levels actors – professional craftsmen and engineers as well as firms and R&D units – learn and develop new knowledge more or less influenced by each other, contributing to development of the actors themselves as well as of the whole system. This analytical approach is an essential part of understanding the processes that occur in industrial districts (and the innovation system) and in the economy. Learning, however, is not the same as innovation, as we will see below.

Knowledge formation and learning are thus core processes in the transformation of economies. Not least the American economists Nelson and Winter (1982), who have written the probably most important book on Evolutionary Theory and Economic Change, have a strong focus on knowledge, skills and competences on individual and organizational level. They were among the first scholars who focused on the role of non-codified or tacit knowledge – a concept originally developed by Polanyi (1967) – in the learning processes of the economy.

The evolutionary perspective of Nelson and Winter focuses on the tension, or balance, between the creation of variety (the invention process) on the one hand; on the selection process on the other and, on the third, on the retention created in the system by the development of routines to keep the system (and production) going. In particular, it is important not to ignore the selection processes, something that is often done by innovation researchers as well as policy units. Most innovations fail, and have to fail, before, during or after their introduction to the market. Nelson and Winter argue that we here find one of the classical failures among innovation researchers: the neglect of the demand side.

Actors (networks, systems) who do not – through learning – have the capability to adapt their routines, products and processes to compete with innovations in their fields of activities (markets) have to close down or restructure their activities. There is, thus, a connection between learning systems and innovative ones (Lundvall, 1994), but they are not identical. Although innovations, following Schumpeter, on the one hand may be characterized as new creative combinations of knowledge which may be recently acquired as well as known since long – and thus has to be (re)learned – they, on the other, must be looked upon as breakaways from ingrained and learned paths; i.e. processes of unlearning. Innovations destroy existing structures and create something new on the ruins of the old. This destruction may be more or less radical: nevertheless, it represents an act of creation – and creative destruction – that is more than just adapting to old routines (Schumpeter, 1942).

Marshall realized that intra-district (local) learning and knowledge transfer from one generation of craftsmen to another, under certain circumstances, could contribute to cementation of old competencies and to an incapacity to
innovate new products or processes. The diligent and quick to learn pupil is far from always the most innovative, and the innovative violinist breaks at least some of the routines learned from the master. In this context. “learning” should not be identified as identical with, or enough for, the variety – selection – retention processes.

This view on innovation as a creative process combining learning and unlearning also challenges the conventional view – still common among policymakers and some parts of academia – on the innovation process as fundamentally linear. Strongly simplified, the linear model of innovation displays an understanding of innovation activities as did they start with basic knowledge (e.g. basic science), followed by more applied knowledge (e.g. applied science), product or service development (e.g. by engineering) and later close-to-market-activities. As argued since long by many innovation researchers this model – strongly connected with the US post WWII report Science the Endless Frontier (Bush, 1945) – ignores, the fact that innovation processes take place over and between different systems levels and domains of knowledge. Innovation is in essence a highly interactive and iterative process, where various knowledge components (e.g. market knowledge) repeatedly come onto the scene, interact, change and shape new understanding (Rickne, 2000). For example, the design elements can be, and often are, substantial and (recent) scientific knowledge may be either present or absent in this creative process (Utterback et al, 2006). Although there are many academic contributions to this discourse (e.g. Faulkner, 1994) the probably most well-known challenge of the linear model has been written by Kline and Rosenberg (1986). Of central importance for their analysis is the stock of knowledge to which all activities – also R&D – contributes and which may be utilized in all phases of the innovation process which in its turn is characterized by significant feedback loops and engineering and design activities.

Nevertheless, is the linear model – with its strong focus on the analytically biased basic sciences – still alive as a blueprint for significant parts of the innovation and technology policy discourse. To some extent this is probably because parts of the academic community has knowledge interests in it, because the linear model is easy to communicate and finally because it is attractive for politicians and policymakers who can argue that money to basic science contributes to the long run development of industry. Naturally, basic science may indeed, and often do, contribute to and is essential to innovative change, but our point here is that there is not a linear or necessarily causal relationship.

Some may argue that there is a part of the IS literature which more or less uncritically adopts and/or basically ends up in the linear model although, as argued here, this mode is just a sub set of all possible forms of innovation. It may, however, also be argued that the IS approach may be used as a
framework for identifying the complexity and network character of different innovation processes.

**The innovation system concept defined**

Let us return to the IS definition given in the introduction: an interrelated structure of institutional – incl. cultural – and actor based condensations in an economic space and which has specific industrial and innovative properties compared to its environment with which it has exchange. The “innovation system framework” may in fact be looked upon as a set of concepts with family resemblance, containing terms with more or less similar meaning. Members of the core family are notions like “national innovation system” (NIS), “regional innovation system” (RIS), “technological (innovation) system” (TS), and “sectorial innovation system” (SIS). We may also include notions like “development block (DB)”, “competence block (CB)”, “industrial district (ID)” and, not least the “diamond” and “cluster” terms once introduced by Michael Porter (1990). Worth mentioning – but not part of the IS family as such – are the so called multi-level perspective (Geels, 2002; 2004), and sociotechnical system concepts such as those presented by Bijker (1995) or Hughes (1992), even though they will not be explicitly discussed in this paper. (see Blomkvist and Johansson in this volume)

Let us then turn to a more detailed presentation and analysis of the different members of the IS family. National innovation system (NIS) was the original concept introduced around 1986/87 first, probably, by the British economist Christopher Freeman (who in fact had used the concept already in 1983, cf. Freeman, 1987, 1983/2004) in his analyses of the Japanese economy. The NIS concept was also – in connection to Freeman – used by the Swedish-Danish economist Bengt Åke Lundvall in his analyses of the competitiveness of small European economies (Freeman and Lundvall, 1988). Around 1990 many texts were published where the NIS concept was used (Dosi et al., 1988; Lundvall, 1992; Nelson, 1993)

It is easy to recognize that the 1980s was an ideal and logical foundation for this kind of theories. By then the post WWII catching-up processes from the

---

24 The IS approach may perhaps not be seen as a formal ‘theory’. Instead, as Edquist (1997) expresses, the IS concept should primarily be understood as a “conceptual framework”.

25 This is in itself a concept originally formulated by Wittgenstein in his Philosophical Investigations (1953/89).

26 This notion, as defined by Carlsson and Stankiewicz (1991) should not be confused with Hughes’ (1992) term ‘technological systems’. Some literature now refers to the type of innovation systems called ‘technological systems’ as ‘technological innovation systems’, to avoid such confusion (see e.g. Hillman et al, 2009).

27 The diamond based analysis and the cluster concept are perhaps the most well-known, being cornerstone in Porter’s analysis of the “competitive advantage of nations” (and regions).
first wave of Asian tigers was maturing: not only Japan but also South Korea, Taiwan, Hong Kong and Singapore challenged the incumbent industrial nations. Several of the newly industrialized countries were not neo-liberal unfettered market economies but rather characterized by strong institutions with high ambitions to favour industrial dynamics. The success stories from these catching-up countries are in sharp contrast to the stagnation and lack of policy that was revealed in the incumbent world (Freeman, 1987; Nelson, 1993; Dertouzos et al, 1988).

There is also a connection between the NIS concept, formulated in the late part of the 80s, and the heritage from the German economist Friedrich List. Already in the first known text on NISs there are significant references to the national production system concept once developed by List in his analyses of the industrial competition between the incumbent and free trade oriented United Kingdom – benefitting from economies of scale and being far ahead on the learning curve – and the emerging German and US catching-up economies (Freeman, 1983/2004).

The 1980s was, in addition, a period characterized by globalization as well as emerging pervasive technologies, like the ICT, which in the extension were assumed to threaten the small countries’ possibilities to stay competitive (Laestadius, 1980; Sharp, 1983; Freeman and Lundvall, 1988). From a northern European perspective there was also the problem of whether small countries, like Norway, Denmark, Sweden and Finland, could maintain competitiveness on country level within an integrated Europe. All in all, this contributed to give the national innovation systems discourse a “small country bias” (Freeman and Lundvall, 1988; Lundvall, 1992).

Somewhat simplified it may be argued that this also is mirrored in the two dominating NIS texts from the early 90s: one European and one American. The comparative approach in the American anthology (Nelson, 1993) is more focused on the analysis of the existence of innovation systems than the European anthology, edited in Aalborg/Denmark (Lundvall, 1992), where most papers focus on the character of the systems. Several of the papers in the American anthology argues for caution as regards the potential to understand, still less to manage, innovation systems (Mowery and Rosenberg, 1993).

The national innovation system obviously fulfilled a need (for academics) in analytically knitting evolutionary innovation theory with institutional theory as well as to connect industrial transformation/dynamics with policy, here illustrated with an influential text by Stan Metcalfe (1995, p. 462-63).

“A national system of innovation is that set of distinct institutions which jointly and individually contribute to the development and diffusion of new technologies and which provides the framework within which governments form and implement policies to influence the innovation process. As such it is a system of interconnected institutions to create, store
and transfer the knowledge, skills and artefacts which define new technologies. The element of nationality follows not only from the domain of technology policy but from elements of shared language and culture which bind the system together, and from the national focus of other policies, laws and regulations which condition the innovative environment. In the operation of national systems, governments play an important part in their support of science generally and in their procurement of technologies to meet the needs of the executive. To define such a system empirically one must locate the boundaries, its component institutions and the ways in which they are linked together."

The NIS concept developed into an organizing metaphor that successively colonized more and more space in international policy documents (OECD, 1997; 1999). Not least in the OECD (1999) document – *Managing Innovation Systems* – national innovation systems appear to be a question primarily of *governance*. By then – in the late 90s – the original NIS concept had developed into a general concept but also bifurcated into a family of varieties (Edquist, 1997). In a world where knowledge formation/learning as well as industrial activity takes place on a global scale it is far from obvious that the essential and systemic condensations are national or even territorial in any real sense. With introducing the concept *sectorial innovation system*, the Italian economists Breschi and Malerba intend to focus on the industrial rather than the territorial dimension (Breschi and Malerba, 1997; Malerba, 2004).

Analogously, although with focus on technology rather than industrial sectors there also emerged a research tradition on *technological systems, or technological innovation systems (TIS)* (Carlsson and Stankiewicz, 1991; 1995). The difference, in short, between these approaches can be explained by the fact that the logic of technological change does not necessarily follow the logic of an industry. ICT, e.g., may penetrate and influence various industries in different ways: Some advanced process industries may, to a certain extent be more influenced by ICT than the IT industry itself! Also biotechnology is often realized in unexpected industries: functional food is, for example, the result of biotechnology that has colonized a small niche within a traditional low-tech industry. Technological cognition processes, in addition, do not necessarily take place within “communities” or “thought worlds” that follow industrial classification or regional or national borders. Although craftsmanship and technological cultures (as once described by Marshall, see below), may develop locally, fundamental technological paradigms and regimes develop globally among researchers and engineers and condition and influence the direction of technological change (Dosi, 1988; Nelson & Winter, 1982; Laestadius, 2000). Today the TIS literature is closely linked to that of socio-technical transition management and a multi-level perspective (Hillman et al, 2011).

---

28 Several analysts within this research tradition changed the “brand name” to *technological innovation systems* (TIS to make it more in line with the other IS-approaches.
This distinction is far from unimportant and has policy implications: if *industries* are targets for industrial policy we have reasons to expect other results compared to if the policy focus is on *technologies*. Not only will there be different receivers of the resources connected to the policy; it will also develop needs for other policy tools and different routines.

The concept *development block* (DB), formulated by the Swedish economist Erik Dahmén (1950), deserves to be mentioned in this context although his block concept is discussed in detail by Laestadius (in this volume). The DB concept contributed to the development of the NIS concept among the Danish economists. Still more has it influenced the technological systems approach (Carlsson, 2000). The development block – which theoretically may be argued to slide between a sectorial and a technological logic – is based on the notion of *structural tensions*, i.e. imbalances occurring due to the very character of the innovation process. There are always – in every period – *necessities* and problems that have to be handled, as well as *opportunities* that may be utilized. These imbalances may develop on the level of technology in the form of bottlenecks versus local breakthroughs, as analysed by Tomas Hughes with the concepts *salients, reverse salients* and *critical problems* (Hughes, 1992). But they may also occur on industry/sector level: the rapidly increasing electricity output due to large scale construction of hydro power plants in Sweden contributed to the establishment of energy consuming TMP pulp and paper plants. The DB concept – which never explicitly was formulated in territorial terms – may be looked upon as a means of making Schumpeterian dynamics empirically useful and adds a structural dimension to Schumpeter’s entrepreneurial focus (see Laestadius, in this volume). Dahmén also stated that development blocks may be used for *ex post* analyses by academics as well as intentionally *ex ante* by industrialists (or policy makers) to create and/or exploit systems imbalances.

**The geography of innovation**

Clearly for the sectorial and technological approaches of IS, innovation processes are not non-territorial, but exists in a geographical reality: The system is seen as global but composed of – and often analysed in terms of – geographically distinct nodes (e.g. a country or a region). In the *regional innovation system* (RIS) approach geography is at the core of the analysis (and knowledge areas of sectors secondary). All the IS concepts, but perhaps best the RIS concept illustrates the difficulties to distinguish a distinct level of territorial analysis in a globalized economy. RIS is an approach with a clear inspiration from the NIS writings and is to large extent based on a similar analytical approaches (Asheim and Gertler, 2005), but also bear resemblance to Porter’s reasoning.

The regional dynamics dimension has been there all the time although branded differently. Specifically, Alfred Marshall’s concept *industrial district*
is highly relevant in this context (Marshall 1890/1990). The Marshallian district was e.g. an important tool for the understanding of the industrial dynamics of northern Italy (Brusco, 1982; Amin and Thrift, 1994).

Also the rapid industrial development in Silicon Valley – which from many aspects differed from the Italian – has contributed to the growing interest for the local and for the territory as well as for analysing the socio-cultural networks and processes which contribute to an innovative atmosphere – or to the absence of such climates (Saxenian, 1994). Silicon Valley – the dynamics of which to a large extent coincidence with, or even was the carrier of, the ICT boom during the 1970s and 80s – has been the victim of often too far reaching generalizations based on one extreme case: not least the interpretation of Silicon Valley as a “science park” has contributed to exaggerated expectations – often connected to policy failures – on how to use policy to create industrial dynamics based on local universities.

The differences between, among other things, the low-tech industrial creativity in some regions on the one hand and the significant clustering of R&D labs in some localities on the other has contributed to a family of RIS concepts. One way to capture these differences is to make a distinction between institutional regional innovation systems (IRIS) and entrepreneurial ones (ERIS) (Asheim and Gertler, 2005). Although this distinction is not necessarily exhaustive – both systems may in fact have a low-tech/high-tech dimension – it serves the purpose to reveal the fact that government, academia, firms and other institutions play different roles in different forms of regional innovation systems. This is, of course, highly relevant for policy institutions: the variety of regional innovation systems may respond differently to various types of policies.

Globalization, it may be argued, paradoxically has increased the interest for the regional dimension in industrial dynamics. The existence of “sticky places in slippery space”, as formulated by Markusen (1996) two decades ago, brings up the question on the role of regional mechanisms – if any – in creating industrial creativity in a totally open economy.

Just to assume that innovativeness grows in a region as a consequence of territorial co-location of firms does not mirror reality in the globalization process of our time (Cooke, 2005). Although the presence of social interaction, trust and local institutions is essential for the development of clusters this does not exclude that also significant non-local knowledge may be necessary to maintain or develop competitiveness (Maskell et al., 1998; Asheim and Herstad, 2003; Cooke et al., 2000; Isaksen, 2005). Regional clusters must develop interfaces with other parts of the world (Rickne, 2000). Interactivity as regards learning and knowledge formation, thus, must not be restricted to geographic (territorial) proximity only (Coenen et al., 2004; Torre and Gilly, 2000; Boschma, 2005; Malmberg and Power, 2005; Gustavsson, 2009).
The complex interrelationship in the regional system is characterized by interdependence on several levels of which the territorial level is just one of many possible (Howells, 1999). Within the framework of these interrelationships – which are more or less open towards the rest of the world – certain activities tend to locate to specific regions (Asheim and Coenen, 2006; Cooke et al., 2000). A regional innovation system is thus not a national innovation system writ small. Even if there are similarities between the NIS and RIS level of analysis – e.g. in the institutional structure – we argue that a significant part of the difference is in the “embeddedness” (Lundvall and Maskell, 2000). In short: the national innovation system is primarily institutional also from a strict point of view: politics, laws, authorities, school system, culture, politics, incentive structure in the economy, etc. The regional system, on the other hand, is primarily social and cultural and contains cooperation, exchange of experiences (transfer of knowledge). The regional level – it is argued – is more based on social relations than on institutions (Granovetter, 1985). However, this is far from always obvious: as will become clear below the distinction between NIS and RIS become blurred when comparing countries of very different size, like China and Sweden.

Even if the world has become “slippery” human activities have to be grounded somewhere on the globe, and in a “globalized” world – with global access to competence localised elsewhere – all kinds of even marginal competitive advantages in a certain location may serve as a mechanism for competitive advantage, and maybe initiate a path dependent process. From these mechanisms, tensions are created between the local and the global. It is, thus no paradox that the interest for territorial analyses increases parallel to the globalization process.

The competitive analyses by Michael Porter also belong to this research territory; his analyses, in fact, span between the national and the regional (Porter, 1990). Although the concept “innovation system” is missing in his magnum opus his analyses show family resemblance with the IS approach. The Porter diamond may be looked upon as a pedagogically attractive formulation of a (national) innovation system although the “government” in the original diamond (Porter, 1990, chpt 3) is given a much more limited role than what European analysts probably would assume. In the more empirical based part of the book the concept cluster – which Porter may be given the credit for having introduced into the analyses of industrial dynamics – has a strong family resemblance with the Marshallian industrial district and the regional innovation system. The clusters a la Porter, which are analysed through the diamond approach, are characterized by competition and rivalry as well as co-operation and networking. As was the case with the NIS concept, the analysis by Porter naturally received its inspiration from the territorially located industrial dynamics which was part of the boom during the 1980s and 1990s to a large reflecting as well the development of the IC technologies and
globalization. Recent research has taken a more nuanced and critical view on the functions – and non-functions – of clusters showing that it sometimes has been too easy to identify and make policy and location decisions to “clusters” (Asheim, Cooke & Martin, 2010).

Over time government has taken a more prominent position in Porter’s research. This is especially the case in the paper *Green and Competitive* where the importance of a high-end domestic demand, created by an advanced environmental policy, for the creation of green industrial clusters and competitiveness is analysed (Porter and van der Linde, 1998). It may be argued that part of their argument misses the point: the transformation pressure discussed did not have its origin in government but in the environmental movement (Cerin, 2004). Institutions are more than governments.

Common for these research approaches is the ambition to identify the combination of locally unique and relatively sticky mechanisms and cultures favouring innovativeness and creativity, i.e. the origin of competitiveness: to have the capability to develop something which does not immediately slip over to other actors, networks or regions.

The implicit assumption in many theories on the mechanisms of dynamic territories where certain attractors make them grow more rapidly than others is that people, often understood as labour, are the agents most inclined or even forced to move. Historically this may well have been the truth. Some years ago, however the Italian-American sociologist Richard Florida in a set of books (2006; 2002/2014) argued that in a world where creativity and human capital is important for competitiveness and professionals are in short supply, high tech firms have to move where the creative class prefer to live. Although Florida’s original conclusion had conceptual and empirical weaknesses (cf. Nuur & Laestadius, 2009) his basic argument is of importance: in a world where human resources are important people do not only adapt to the labour market but create the conditions for industrial life.

More sophisticated discourses on the contextual ground for knowledge formation (and learning) do, however, take us away from a narrow interpretation of the territory. Introducing the concept *spaces of knowledge* – reminding us of the concept *economic spaces* introduced by Perroux 1950 already – the geography of knowledge is no longer fettered to the territory (Amin and Cohendet, 2004, chpt. 5). This also takes us to the approaches common in modern sociology and management theory. Specifically, a significant part of the knowledge management theory is focused on how to develop and maintain learning and knowledge creation on corporate level in global firms and thus basically independent from territorial aspects (von Krogh et al, 2000).
Creative and innovative processes can nowadays be organized on global level as well as on the regional. Multinationals that traditionally have had their R&D units in Europe or the US (Pavitt and Patel, 1991) now relocate their R&D activities. Within some scientific and technological areas there are also research communities strongly connected in networks with low physical proximity but maintaining and developing professional relations on distance. Phenomena like these do change over time: mechanisms which contributed to proximity some decades ago – e.g. in the industrial districts of Alfred Marshall – may be less relevant today. And information technology development may create potentials for communication that, at least partly, may substitute for direct human interaction. How far this globalization of knowledge formation has gone and whether there are limits to this process is the topic for much of present day research. It may be assumed that innovation systems are changing in character due to internationalization (Carlsson, 2006).

What constitutes a region or a regional system is thus far from evident. We can imagine a set of relations which all have different geographical anchoring – and distributed over different geographical territories – but still being important for the system in question. The condensations of the different relations may be interpreted as layers of different systems added to each other; this set of layers may reveal a certain territorial structure in the form of a set of nodes (hubs) connected through a set of network relations, together constituting the economic space. The non-territorial systems we discussed above may thus be included in this model.

One way to conceptually handle this variety of relations is to introduce the concept functional regions (Hallin, 2005). Theoretically we thus get a set of regions each of which is defined from a certain identified function and not necessarily identical with a classical geographical or administrative structure. A functional region from an innovation point of view must not coincide with the regional labour market.

It may be argued whether functional regions – strictly speaking – are territorial or not; that is of no importance for our analysis here. Of importance, however, is that (innovation) system borders no longer – if they ever could – can be assumed to coincide with those administrative borders created by, and conditioning, the activities of policy makers.

Enlarging the conceptual ambition to the global level the innovation systems approach faces a lot of problems. The meaning of concepts like global or continental innovation systems – as introduced by Freeman (2002) – is far from obvious. This problem was revealed already in the early analysis by Mowery and Rosenberg (1993) relating to the U.S. innovation system. Still more do we face the problem of analysing the industrial transformation in China which has – with 1400 million inhabitants – during three decades (since 1978/79) shown a unique industrial transformation; longer and more
rapid than any other country in history. But hitherto, most of this dynamics has taken place in a segment of Chinese society covering approximately 300-400 million people of which the majority live in three large regions (Pearl River Delta, Yangtse River Region and Bo Hai Rim) and in another half a dozen big urban areas with populations up to the magnitude of 30 million (e.g. the Chonquing area). But in these fast-transforming regions there are large segments of the population that are outside, or even losers in, this dynamic. It is difficult to imagine that the NIS and RIS concepts, as used in the European way, could have a similar meaning when applied on countries like China. Every dynamic “region” in China is larger than the biggest European countries and the big industrial cities have a larger population than a typical small or medium sized European country. In addition, it should be observed that much of this transformation – or catching-up process – more has the character of classic growth – i.e. more roads, canals, railways, cars and houses – rather than innovations as usually defined (Lundvall, 2006; Laestadius et al., 2008). In short: we are here probably approaching the limits of the innovation system concept.

Components and critiques

Within the IS families a set of core concepts are embraced. We will here discuss and problematize some of these central aspects, and point to the need for further clarification, specification and empirical investigation of the IS framework. A first core feature is the explicit identification of the importance of institutions in conditioning economic processes. Secondly, we discuss the general understanding of the system concept, particularly focusing on openness and intentions. Thirdly, we scrutinize the identification of systems borders that may be related to geography – but also to other forms of proximity (knowledge/technology). Fourthly we focus on the problem on to what extent a “structure” is a feasible analytical concept when analysing the dynamics of industrial and technical change. Finally, we turn to the governance question. The aim, of course, with IS related analysis is to obtain results that can be used for policy and management. To what extent can ISs be governed, or even created – and how?

The role of institutions

In the definition of ISs used above there is an implicit although fundamental criticism of the core message in standard economics: the notion of pure atomistic equilibrium markets – i.e. characterized of all the requirements for full competition – is not a good approximation of real economies. These are always embedded in an institutional and cultural web conditioning activities in the long run as well as in the short. Our focus on the role of institutions – contributing to the creation of the system as well as being created by it –
follows a long tradition among economists and industrial analysts (Hodgson, 1998; North, 1990). The IS approach may be looked upon as a holistic attempt to understand the institutional web of which the classical economic agents, “firms” and “labour” are just two elements of many.

The role of institutions has been a core research topic for economists for a century at least. Among the most well-known among international classics are Thorstein Weblen, Douglas North, Ronald Coase and Gunnar Myrdal. Among Swedish economists, not so well known in this area, we also find Johan Åkerman and Ingvar Svennilsson (Hodgson, 1998). For some social scientists outside the economics discipline – e.g. sociologists and political scientists – the important role institutions take is by definition obvious. What is the book by Putnam on the Working Democracy if not an institutional analysis of the north and south Italian ‘innovation systems’ (Putnam, 1996)?

Understanding innovation systems as an institutional and relational condensation is thus essential. That leads us, to the meaning of the “institution” concept. Here we follow conventional social science theory using a very broad definition for the concept where it is seen as “sets of common habits, routines, established practices or rules which regulate the relations and interactions between individuals and groups” (Edquist and Johnson, 1995, p. 9). According to this view, an institution is every social construct – be it intentional or not – regulating social interaction. Thus, an institution is not a government authority only; neither is it primarily a hospital, a home for elderly or an opera – although these three categories may all be included in the institution concept. Institutions may be acting on the market – patent authorities e.g. charge applicants – but their importance is primarily not in selling and buying but conditioning and influencing the behaviour of societies.

Accepted business behaviour, the working moral, patent systems, property laws and registration systems, tax laws, school systems, entrepreneurial spirit and prevailing views on gender, nature, equality and religion all belong to the set of institutions which contributes to the embeddedness which the IS approach intends to capture. Obviously there is variety as regards institutional embeddedness, and thereby there is also variety between IS as regards the impact certain institutions may have on the behaviour of actors. IS researchers typically have the ambition to examine differences and similarities between different innovation systems, maybe also to point to their perceived or revealed weaknesses or strengths analysed from a certain perspective. While such analyses could theoretically contain a very large number of institutions, they are in practice often focused on a rather limited set. For example, one

---

29 In contrast, organizations are “formal structures with an explicit purpose” (Edquist and Johnson, 1995, p. 11). This means that we do not use the term institutions when we mean the organization as such. For example, a university may be referred to both as an institution influencing societal interactions, and as a specific organization.
may scrutinize what role different countries’ institutions for property rights have for inventiveness. However, while the analysis of cultural norms and behaviour may be just as relevant, such institutions have more seldom been centred in the IS literature. In part this may be due to the knowledge background of the community of IS scholars, and also on the theories and methodologies developed to measure such phenomena. For example, do we really have the tools to evaluate whether the welfare system, the role of ethnical minorities, the gender attitudes or the R&D policy is most important for the innovation capacity of a certain region?

The systems concept

As we are dealing with a “system” we need to first of all discuss what this implies. In general, systems are made up of components, attributes and relationships. The components are institutions, actors and artefacts, and attributes are the properties of the components. Relationships are the links between these components: Market as well as non-market links. Feedback loops and interaction provide the dynamics of the system. As Carlson et al. (2002, p. 234) put it: ‘One result of interaction (feedback) among actors is that capabilities shift and grow over time, and therefore, the system configuration also changes’. Also, the system’s propensity for to be both robust and flexible, and its ability to both induce and handle change are crucial. Another way to define a system is to see it as “a group of components (devices, objects or agents) serving a common purpose, i.e. working towards a common objective or overall function” (Bergek et al., 2008, p. 408). In this context, many researchers agree that an innovation system have the overall function of developing, diffusing and utilizing innovations.

Already classical systems theory struggled with how to handle the transformation of the systems approach from the natural world (e.g. physics or biology) to the social (Emery, 1969 and Ingelstam, 2002). An important step – especially interesting due to the evolutionary foundation of innovation theory – taken by von Bertalanffy (1950, in Emery, 1969) is his distinction between the equilibrium (closed) systems of physics and the steady state, on the one hand, and open biological systems on the other. Similar to biological systems, social systems are open which makes it possible for them to develop and to maintain or even change their systems structure while transforming their parts.30 Clearly, systems can be more or less open (closed) towards their context. An extreme interpretation with a “totally open” system renders the concept without meaning as one cannot discriminate it from its context. One way to come to grips with the degree of openness is to specify the types of elements that are to be included in the system. In the case of innovation systems literature, one often includes the relevant institutions, knowledge areas, artefacts, actors, networks and the associated knowledge flows. We will

30 Note that the biological analogues should not be extended in all directions.
come back to this issue of delineating the system and deciding upon relevant elements in a more detailed way below.

Another important distinction of the system concept – relating to how the IS concept is used by various theorists and practitioners – is whether or not the system is seen as an analytical construct or as a system that exists in reality. Phrased differently, one may use the IS concept to describe/analyse an empirically observed set of interconnected actors, networks and institutions. The IS framework thus becomes a model for us to better describe and explain identified structures and connections. With this use of the IS concept, one may sketch a potential system that could emerge but that doesn’t exist in practice, or where links or functions are not in place (Bergek et al., 2008).

Assuming that we have identified the existence of innovation systems we may make a distinction whether they are self-organized or constructed by intention. Whether certain innovation systems are social constructs that contain human actors and organizations that have intentions – is a typical research question. Actors may have intentions for their activities (firms are e.g. often assumed to maximize profit and individuals may have certain goals for action) and these intentions may include that some functions of the system as such needs to be strengthened. In this way, actors such as firms may actively work towards supporting (their view of) a well-functioning system (Nilsson et al., 2012). Rarely, however, firms take an overall responsibility for all functions of a system, but are more likely to focus on particular sub-functions. Moreover, policy or bridging organizations may have the intention to construct a system with certain properties. In this quest such organizations may strive to create a system from scratch or to strengthen an existing system. In order to achieve that they may be engaged in the construction and transformation of institutions which are parts of the system.

Having said this about actors’ intentions, our broad definition of innovation systems does not as a prerequisite assume that innovation systems are the result of intentional systems building. In fact, while in some cases collective and coordinated action may be the case, it seems to be more common that innovation systems emerge through the aggregated effect of uncoordinated more or less intentional actions over long time periods. From an analytical point of view such systems may be looked upon as self-organizing where intentions, entrepreneurial culture, etc. may be more or less formed endogenously. As Bergek et al. (2008, p. 408) put it: “Actors do not necessarily share the same goal, and even if they do, they do not have to be working together consciously towards it (although some may be). Indeed, conflicts and tensions are part and parcel of the dynamics of innovation systems.”

When scrutinizing innovation systems, the borders may of course be defined narrowly enough to allow for intentionally constructed systems. Not the least
has this “voluntaristic” approach been popular among policy makers with intentions to create innovation systems.

**System borders**

In our definition of innovation systems above we have situated the IS to an *economic space*, i.e. the domain where all activities of an economy take place. Inspired by the French economist François Perroux (1950), who originally introduced the concept economic space, we choose not to *a priori* restrict our definition to a territorial – *genomic* – space. The possible territorial anchoring of an innovation system thus becomes an empirical question unless we do not explicitly restrict ourselves to geographical proximity. Consequently, geographically defined systems – like regional ones – thus become a sub-set of the large family of IS. Similarly, national innovation systems become one of many possible interpretations of an IS although – as we will come back to below – the NIS concept is the original (Lundvall, 1992, Nelson, 1993).

One may argue that geographical borders become increasingly meaningless when speaking about innovation processes. In this epoch of globalization: what is left of territorially defined innovation systems when not only industrial production but also knowledge formation takes place on a global scale? Is, as formulated by the American journalist Thomas Friedman (2005), the world in fact flat, i.e. can everything more or less take place everywhere with zero friction and instant connection? And does product and service innovation require ever more knowledge domains to be integrated, rendering the analysis of specific technologies or sectors less meaningful?

While there may be more than a grain of truth in such statements, innovation processes are nevertheless characterized by space as well as by the specificity and logics of knowledge domains, and the IS concept tries to capture such features not only statically but also over time. This means that each IS analysis needs to define the borders of the system in question. By default, each such attempt to delineate a system must carry its inaccuracies: There are no absolute borders to an open system but only the more or less arbitrary ones set for the purpose of analysis. In this way an IS is an analytical construct, and not an absolute reality. In general, there are three dimensions that are used to delineate any specific innovation system: geography, knowledge domain and time. For each of these dimensions’ issues of level of analysis become important. Depending on which of these starting points the analyst chooses – geography, knowledge domain and/or time – he or she will see different systems. Not the same set of components (institutions, actors, artefacts) or relations will be included in the structural description of the IS, nor will necessarily identical attributes of the components come to focus. This implies that the choice of perspective from which to look at the innovation processes will be crucial for the findings. For example, being interested in the emergence of environmental friendly alternatives to the combustion engine, one may
partly stress dissimilar aspects having chosen different geographical scopes and levels of analysis (e.g. the RIS of Western Sweden, the NIS of Sweden, or the IS of EU). Clearly, the question at hand is what will guide the choice of borders for the system (Carlsson et al., 2002).

What may by some be seen as a weakness of the IS framework – that there are no clear guidelines as how to discriminate the system and its parts from its environment – may also be considered the strength of the framework giving it its flexibility. As we do not a priori (e.g. by definition) know what phenomena or what components to include in the various innovation systems, the choice of delineation will be essential and has to be well informed.

While these issues of border setting are decisive – and far from easy – they also will lead to further questions relating to the level of analysis and measurement matters. For example, should we only count condensations of (which?) firms, or should we include (which?) individuals (and in what of their roles)? What level of analysis should one choose for the demarcation of a technological innovation system? How do we count regional industrial service offices and technical universities? And how do we include change and transformation in the systems analysis? These questions have been dealt with in several methodological texts related to the IS concept (Carlsson et al., 2002; Bergek et al., 2008, Hillman et al, 2011; Magnusson and Rickne, 2012), but more is yet to be done in this vein of research.

**Dynamics**

A common set of critique towards the IS approach says that in spite of claiming to evolutionary aspects of innovation there is a risk of rather static and structural analyses. There are several reasons for that critique. One is the relative lack of tools to capture dynamics in a structure as well as in the function of a system. That problem has since long, been in focus among systems analysts (cf. texts in Emery, 1969). In short the problematique may be described as the relation between structure and function in defining the system. As regards the structure – functional dimension in social systems this was early identified by e.g. Merton (1949).31 Is the system basically defined from its structure (its components) and if so, has the system changed – or is it another system – if components disappear and/or transform? Or, is the system basically defined from a set of essential functions or purposes with the implication that the system is basically the same irrespective of structural change as long as the basic functions are there? And more fundamentally:

---

31 And this is particularly problematic to combine with an evolutionary perspective, a problem with which both natural scientists (biologists) and social scientists have struggled. The usefulness of a functions concept differs between physical systems and biological ones as well as between natural science related systems on the one hand and social systems on the other (cf. Nagel, 1956 in Emery, 1969).
What are the functions of the system, and how can these be confined? This is, of course, also relevant for the analysis of innovation systems.

Another problem for systems researchers in general is how to measure the performance of a specific system as well as the relative importance of its specific attributes. A third and related problem is how to obtain relevant data for comparative analyses between systems that largely deviate from another.

In social sciences this balance between structure and function of systems – and the epistemological consequences of that – has since long been in focus for many disciplines: anthropologists (Radcliffe-Brown, 1952), sociologists (Parsons and Smelser, 1956; Merton, 1949) and political scientists (Almond, 1960). The general problem is well illustrated with a case from political science: in his classical paper from 1960 Almond identifies the political systems more or less completely from its functions of which he identifies four essential input, and three essential output, functions.

During the last decade, several authors have addressed this critique by developing a functional approach to IS dynamics (Galli and Teubal, 1997; Johnson, 1998; 2001; Rickne 2000; Johnson and Jacobsson, 2001; Liu and White, 2001; Bergek, 2002; Bergek and Jacobsson, 2003; Carlsson and Jacobsson, 2004; Bergek et al., 2008; Hekkert et al., 2007). In essence, these researchers argue that each IS – basically their focus is on technological innovation systems, TIS – can be portrayed by a number of functions describing what actually ‘happens’ in the system. The set of functions identified is from the beginning empirically derived. Complemented by findings from e.g. sociology, organizational theory and political science the functions arrived at have a broad resonance in the literature. The approach describes how to assess the “functional pattern” of the IS: This implies to determine how the key processes currently work in any specific IS. Bergek et al. (2008, p. 414) stress that the “functional pattern of a TIS is likely to differ from that of other TISs and is also likely to change over time. Thus, the concept should not be interpreted as implying that the pattern is either repeated or optimal.”.

The Swedish “TIS functionalists” – suggest that the systems are based on the following functions (Bergek et al., 2008):

- knowledge development and diffusion
- influence on the direction of research
- entrepreneurial experimentation
- market formation
- legitimation
- resource mobilization
- development of positive externalities
This list can be argued and the definitions questioned: do the functions constituting TIS differ from those of the NIS, the SIS or the RIS? Or, do the set of functions differ over countries or time? While the functional approach by no means solves the entire dilemma of measuring dynamics or performance it is nevertheless a valuable step along that road. The epistemological question on how to identify the functions of a system is indeed not trivial. It may be argued – and has in fact also been – that such functionalistic approaches run the risk that the list of functions gets an ad hoc character, that it can be extended or reorganized with no obvious theoretical ground. How do we validate that the 4+3 Almond functions make the tool box complete?, or in general terms: is my list of essential functions more reliable than yours? Nevertheless this functional approach opens for research activities to get answers on these questions.

This dilemma can be illustrated by the following “constructed” illustrations: if an IS is assumed to be to less (highly) entrepreneurial – how do we obtain knowledge of whether that is a consequence of the school system, the tax system, the (lack of) protestant ethic, the welfare state or the innovation policy? And if a NIS produces a lot of R&D results but is not outstanding in GDP- development – how do we know what functions – if any – to blame for that. In short how do we obtain scientific results that differ significantly from pure prejudices?

**Governance**

Moreover, the functional approach among IS analysts may be looked upon as one answer (among others) to the “governance dilemma”. This impasse relates to the fact that there are many actors who individually or jointly endeavour to put various schemes into practice (let us call these governance arrangements) with the aim to influence innovation input, its processes and its outcome: We call this process innovation governance. The dilemma as such lies in the difficulties in guiding or controlling such complex and ever changing innovation processes, where so many actors and knowledge areas are involved. The governance concept is wider than public policy: We include all actions and actors involved in guiding or directing the process at hand. This implies that public policy actors may or may not be crucial actors: There is nothing in the governance approach admitting a special role for public policy. Instead, various types of actors may take on different responsibility depending on e.g. technological or sectorial area, geographical conditions or current trends.

---

32 The Almond (1960) input functions are 1) political adaption and recruitment; 2) articulation of interest; 3) aggregation of interest, and 4) political communication. The output functions are a) rule making; b) rule application, and c) rule adjudication.
Not the least may a functional approach contribute in relation to the globalization problematique: sorting out the various functions to be filled within an IS and making them comparable over space and time. Globalization challenges, for example, the balance between the regional, national and supra-national (Rickne, Laestadius & Etzkowitz, 2012). Which structural components and which functions can be expected to be served at which spatial level? Is it reasonable to anticipate clusters to emerge within e.g. biomedicine at a multitude of locations around Europe: And what does this imply for national policy actors striving to support such clusters? Another governance issue related to globalization is what types of actors take what roles in pushing the regional, national or supra-national agenda.

All in all: governance issues are central when it comes to innovation systems. However, even given the possibilities of the functional approach, critiques assert that while the institutional set up is used to delimit the systems considered, institutional changes and influences, as well as governance aspects, have more seldom been fully captured by IS analyses. In fact, current theoretical development has not coherently shown how to handle these issues (Jordan, 2008). In particular, there is a lack of integration with established governance approaches (Treib et al., 2007; Newell et al., 2008) and the IS framework. A recent attempt along this line was presented by Hillman et al. (2009), merging IS theories with the multi-level perspective (Geels, 2002; 2004) and governance understanding to form a comprehensive framework.

The governance issue can also be read in the more policy focused IS documents which, in particular, have been published by the OECD (OECD, 1999; 2002; 2005a-c). Already in the early OECD work on innovation systems it was felt among policy makers that the institutional set up for policy was not adequate for the transformations in innovation systems that were taking place, in particular the growing importance of innovations and the decline of the linear model (OECD, 1999; 2002).

Later OECD reports have extended that analysis and, based on scanning state of the art practices in several – primarily small – OECD countries identified the tensions in the innovation systems and the potentials for policy formulation and integration (OECD, 2005 a-c). Explicitly the growing importance of sustainability and information society policy is identified as new areas to consider in the reorientation of IS governance. Among the approaches intended to handle the transformative challenges related to sustainability, and spanning over several layers of innovation systems, we above mentioned the multilevel analyses by Frank Geels and colleagues (Geels, 2002; 2004 & 2015).
Managerial and policy implications

The aim of this paper was to enhance our understanding of the theoretical basis on which innovation policy in several countries rests. While we did not aspire to synthesize or develop new theory, our intention was to remind those who use the innovation systems tool box that it is important to be aware of the shortcomings and problems of this analytical framework as well as its advantages.

Modern innovation theories rest on an evolutionary view where learning processes and path dependencies are key, and where variety creation, selection and retention mechanisms are central and where equilibrium is imaginary. In fact, this non-existence of equilibrium creates the tensions which constitute the engine of innovation and change. The innovation system concepts in its various variants, is a means for framing all these processes in their context. The system approaches have, we argue, compared to earlier equilibrium based theories, come closer to the balance between structure and function in the innovation landscape. It seems that one important factor in making the frameworks useful for governance of transformation and innovation processes has been the continuous interaction between policy, practice and theory.

Nevertheless, there are certainly several flaws in how the frameworks manage to depict a complex reality: openness and intentions in systems as well as delineation of systems borders remains problematic tasks for the analyst. Another area is to what extent innovation systems can be created intentionally or can/should be the main framework for innovation policy. Our review has aspired to highlight such difficulties, thereby giving analysts, governing actors and managers a better ground on which to stand.

33 For Sweden these frameworks are ways to address issues of industrial dynamics. Indeed, while several countries and supra-national units as OECD or the EU have utilized the innovation system approaches to varying degrees during the last decades, Sweden is perhaps at the extreme end of this spectrum in that the IS concept has been allowed to set the agenda for national and regional policies and program details, and even to name a governmental organization, Vinnova.
References


5: Location as a Matrix of Competition
Cali Nuur

Why is Microsoft located in Seattle, Washington in the USA? Why have places like Silicon Valley, Northern Italy, and Baden-Württemberg in Germany and close to home, Gnosjö become synonymous with industrial competitiveness? What are the mechanisms that influence industrial location? Why are some regions rich while others poor? Why are some industries attracted to locate in a specific area? Why are there regional variations in terms of competitiveness? What are the factors that influence the concentration and dispersion of production? Or to put it simply what is in location/place?

For many decades, if not centuries, the importance of location has been the subject of academic, regional development and industrial policies’ discussions. In the academic debate, a number of concepts have been developed which have also become pervasive in policy circles to enhance regional competitiveness. These include, but not limited to, concepts such as industrial districts, clusters Growth poles, Technopolis, science parks, creative cities, regional innovation systems, just to name a few. With slight variations, the underlying theme of these concepts is that what takes place in a location affects the competitiveness of firms, people and have policy implications. The aim of this chapter is to provide an overview of some of the key concepts relating to location.

The earlier literature

Long before the dawn of modern-day means of transportations and communications, e.g. cars, jets and Information and Communication technology (ICT), the German economist Johan Heinrich von Thünen

1780-1850

provided a rational for the location of economic activities. In what has become known as Land Use Theory, he used a circular model and illustrated the relationship between supply and demand. Others such as Lösch (1954) and Christaller (1966) also developed the so called Central Place Theory which describes urban patterns of location. Also Hotelling (1929; see Hover, 1948, for an overview) has provided theoretical frameworks in regards

to spatial competition. In many of these models, contemporary economies of location such as knowledge formation, externalities and relationship building were ignored as the models were based on assumptions and postulations that were based on achieving equilibria.

For example, Christaller (1966) assumed similar spatial distribution of households, while Hotelling (1929) assumed spatial distribution of price and elasticity (see Parr; 2002; McCann; 2002 and Laestadius; 1992 for further discussions). An important scholar in this context is Alfred Weber (1909) who provided a linear model – a least cost approach- of industrial location that considered transportation costs as an important element in the particular firm’s choice of location. A key conjecture of Weber’s model was that the particular firm pursued minimisation of production, labour and transportation costs when deciding the location of the firm. In addition, the assumption was that the raw materials that the particular firm needed to turn into finished goods were produced in the same vicinity, while lighter materials were produced near the source of raw materials. Furthermore, the assumption was that demand was local. Because production inputs, including labour and raw materials, were assumed to be available in abundance and ubiquitous, and the demand market was assumed to be geographically concentrated nearby the single most important determinant of location was the minimisation of transportation costs. Walker (1989) argues that since the model was based on the conditions of a single plant choosing an optimum location for its operation it did not provide the reasons for industrial agglomeration and the external linkages between the particular firm and actors that provided auxiliary resources.

**Location and development**

In the period following the second world war, there emerged a literature on the relationship between economic development and the location of economic activities. This literature was in response to criticism regarding the shortcomings of the models described earlier and analysed the impact of the location of industrial activities on the region/localities/countries.

In addition, there emerged another set of literature which discussed the mechanisms behind the emergence of the *core and periphery* of economic development. Some of this literature moved the level of analysis to developing as well as nearly independent countries and discussed the underlying reason for un(der)development 35 (see, e.g. Furtado, 1964). Influenced and adopting a Marxist approach the key message was that spatial development – mainly at the country level- is a question of the relationship between centre (old colonial powers) and the peripheral (newly independent countries) where the core dictated the terms and the periphery

---

35 This is worth a chapter of its own

At the same time, even in developed countries new concepts and perspectives that offered new insights and perspectives on spatial competition begun to emerge. For instance, the French economist, Pérroux (1950) objected to the static and abstract models that ostensibly explained spatial competition and proposed the notion of an “economic space” that was based on linkages and relationship. Lausén (1969, p 21) captures the dissatisfaction with previous scholars who he claimed treated the geographic space as “a passive rigid container that conditions the dynamic evolution of economic forces”.

Perroux approach was conceptualised as growth poles (pôle de croissance) and became the point of departure for regional development economists including Thomas (1969), Lausén (1969) and Hansen (1972) to name but a few. There also emerged another stream of literature on the regional dimension of economic development. For instance, Walker (1989) asserts that the presence of a lead industry in a given territory is likely to make the whole region/nation prosper and develop since the successful industry depends on local inputs. Likewise, the same development elsewhere could indeed diminish production in another region (Pletsch 1982).

Another scholar worth mentioning in this context is Walter Isard (1956, 1959 and 1960) who pioneered the field of regional science and made a case for multidisciplinary approach, mainly the inclusion of all aspects of the social sciences, in analysing regional competitiveness. By and large, the birth of regional science has its origin in the criticism on the abstract models, which assumed space as homogenous and available to every actor. These failed to explain why different locations favoured or disfavoured the growth and development of a particular industry. Soon studies were commissioned and concepts developed which viewed the territory a dynamic entity has the ability to influence the decline and growth of industry.

In summary, the conceptualisation of space and subsequent introduction of growth poles theory attempted to answer the questions of: What factors play a role in the economic development of a region; what are the features of successful industries; what mechanisms contribute to the evolution of

---

36 There is a myriad of literature on the impact of colonialism in ensuring the evolution and sustentation of the periphery. Some of these have taken a Marxist approach where independencia was key theoretical point of departure (see, Frank, 1969; Furtado, 1964; Emmanuel, 1977 and Arighi, 1977) while others took a somewhat structural approach (see, e.g. Prebrish 1956). Admittedly, I am name dropping as there is more to this literature stream than merely a critique of orthodox economic theory.
peripheral regions? And how do clusters of innovative firms emerge? Pérroux asserted that territory had the ability to become a stimulus for the general economic condition of a region or a nation by inducing propulsive industries\textsuperscript{37,38} and lead firms (Lever 1980, Tomas 1969, Lasuén 1969). For Pérroux and other development oriented economists, the notion of development is about the birth and death of industries at a particular place i.e. industrial dynamics. Old industries die because product and process innovations make previous products redundant and open up new areas of production.

It might be argued that Pérroux’s ideas were related to the creative destruction approach adopted by Schumpeter (1943), who contended that economic development is affected by factors such as technology and the diffusion of technology as well as endogenous combinations such as the introduction of new products that the consumer has not tried before or a qualitative extension of an available product, the introduction of new methods of production or marketing method, the exploitation of new markets and the organisation of a new industry. However, it is worth noting that Schumpeter did not have the geographical dimension as his point of departure. Perroux’s growth poles theory recognizes the ability of technology to shape territorial dynamics and become a driver of economic development and thus an important factor that is exogenous to the traditional comparative-based advantages of land, labour and capital as the basic prerequisites to growth. Lasuén (1969) suggests that growth poles theory relates to Schumpeter’s theory of innovation in that is also discusses new process and new products that affect the structure and composition of industries in a given location. The geographic location that housed old industries become peripheral while the location of new industries become core regions and start to develop. In other words, innovations contribute to the territory undergoing the same kind of dynamics as that of the individual firm. To argue the similarity of product and geographic space in economic development, Pérroux (1950), who obviously was influenced by the Schumpeterian approach, discusses the features of leading industry, industrial complexes, sectoral and spatial relationships between dominant industrial location and the industry, and the relationship between industrial clustering and the general development of the nation.

As indicated earlier, Pérroux’s approach came at a time when several other social scientists discussed the notions of underdevelopment. Apart from sociologists and political scientists discussed earlier, development economists were also conceptually engaged in the discourse. For example, Albert


\textsuperscript{38} Lever (1980) defines propulsive industry as firms which are “significantly linked by flows of goods with a large number of other industries and which are therefore likely to transmit growth more widely” Lever (1980, p.501)
Hirschman (1958, 1967) in studies done in peripheral regions in the south of Europe and in South America, developed the concepts, *unbalanced growth* as well *polarisation and trickling down*\(^{39}\) to describe how growth in a region is affected by the evolution of new growth poles as a result of a dominant/lead industry. According to Hirschman (1967), the development of a lead industry induces a process of innovations, which positively affect the general progress of the region. The emergence of growth poles as a result of a propulsive industry could lead to the positive effects of polarisation, but it could also have a negative impact in terms of a trickling-down effect on other areas leading to uneven development as core areas attract manpower, capital and public infrastructure from surrounding areas while peripheral areas undergo a process of decay (Amos Jr 1988).

The growth poles as a strategy of regional development gained prominence in the 1960s and early 1970s as governments saw them as an opportunity to influence regional development (Amos Jr. 1988, Polése 1999). These post-war decades were characterised by an economic reconstruction of industry and the subsequent boom of the economies of Europe and North America. In academia, factors such as the availability of statistics, a process of statistically significant urbanisation/migration in some regions and differences in population distributions contributed to the usage of terms such as “disparities” and “lagging”. To reduce regional economic polarisations and to ensure the development of peripheral regions, many governments established departments with the task of pursuing and stimulating regional development policies that could contribute to even development. Through diverse forms of incentives, such as subsidies, investment in infrastructure as well as channelling public funds into lagging regions, governments hoped to encourage regional parities (Polése 1999).

**Industrial districts**

The political economist Alfred Marshall who also wrote the influential book “principles of economics” and largely accredited with bringing the field economics in the realm of science is the sage of the literature on industrial districts. In addition to describing the propulsive ability of the territory, Marshall (1890/1920) laid the foundation for the reasons as to why there are *industrial agglomerations*. In the chapter Industrial organization – the concentration of specialized industries in particular localities, Marshall discusses the advantages that geography could offer to economic actors following empirical observations made on the districts surrounding Lancashire and Sheffield (Laestadius 1999).

\(^{39}\) This should not be confused with the politically laden notion of "Trickle-down economics", also known as "trickle-down theory" which is basically the notion that as wealth increases for the rich, it will result in spillover to the poor.
Marshall (1890/1920) argued that concentration of economic activities resulted from historical circumstances, including factor endowments, local demand and the geographic conditions of the area. Nevertheless, perhaps his chief contribution was in revealing the relationship between the internal capabilities of individual firms and external mechanisms that proximity facilitated. According to Marshall, firms develop internal economies, which allow them to either cut production costs or increase sales. To a certain degree, accessing and sharing resources with other economic players in the same geographic proximity is the result of what takes place outside the firms' borders. In this respect, Marshall played a central role in the theory of economic organisation (Laestadius 1992).

Marshall described economies of location and in so doing laid the foundation for understanding the externalities that firms can derive outside its borders. For the sake of simplicity, economic geographers group agglomeration economies into two economies: urbanisation economies and localisation economies. The former emerge when people and economic activities locate near to each other in order to benefit from proximity. These are often external to the area, but internal to the firms in the area as they gain economies of scale in production (Malmberg 1998). On the other hand, localisation economies are firm specific and emerge because of geographic proximity, which allows the firms to have collective benefits that are unavailable to firms in other areas.

According to Malmberg (1998), both urbanisation and location economies are largely influenced by the same set of factors and develop from shared infrastructure. Physical proximity allows the companies to produce at a lower cost as transportation, thus transaction costs are reduced, and they can benefit from cheaper provisions that arise from shared accessibility to harbours, roads, etc. Furthermore, it facilitates the flow of information and allows the companies to share technical, education and labour market infrastructure.

It is worth noting that in addition to the studies on the relationship between territory and competition, Marshall also discussed various aspects of economic development, including the role of politics and the division of labour (Laestadius 1992). As regards location, Marshall (1920) offered three major advantages that geographic proximity provides to people and firms: creation of a labour market, external economies of scale, and learning opportunities.

First, geographic proximity facilitates the creation of a specialised labour market as it attracts both employees and employers. On the one hand, a person seeking employment is inclined to move to a place where he can get a job and, on the other hand, employers are inclined to settle anywhere where it is easy to select and recruit personnel. This approach is often referred to as localisation economies (Malmberg 1998).
The second important advantage of geographic proximity is the ability to offer external \textit{economies of scale}. Concentration of business activities creates external economies through vertical relationships, such as sub-contracting agreements between firms in the value chain, and vertical co-operation for the firms that produce the same kind of goods (Marshall 1920). At the same time, it gives the concentrated firms advantages in terms of lower costs over firms in other territories.

According to Marshall, the third major advantage of geographic proximity is the ability of geography to enhance face-to-face interactions between economic actors. Asheim (2000) suggests Marshall’s main contribution to social science is in revealing proximity as a factor that enhances the creation of learning opportunities. Through relationship ties and the facilitations of information exchanges between different actors, geographic proximity fosters learning and innovation that provide the necessary resources for competitiveness (Malmberg et al, 1996).

According to Malmberg et al (1996, p 90) the geographic proximity allows for tacit knowledge exchange because this kind of knowledge “does not reside in blue prints and formulae, but is based on personal skills and operational procedures which do not lend themselves to be presented and defined in either language or writing”. This aspect of learning activities is not only developing the competitive strengths of the existing firms, but also of enhancing the growth of new firms was a key aspect of location.

Marshallian dynamics of location, especially with respect to relationship and knowledge spillovers as an important dimension in understanding spatial competition – in particular in regional development policy circles- were confined to bookshelves until Piore and Sabel entered the scene in the mid-1980s. In their much-acclaimed book “The industrial divide” (1984), they discuss the emergence of two major paradigms supposed to have shaped economic development in post-war Europe and the United States. First, they analysed how mass production technology at the turn of the twentieth century ended the era of handicraft production and gave rise to the Fordist era in the USA and parts of Western Europe. Secondly, they describe the stagnation of the world economy in the 1970s and the arrival of new flexible production methods, in which manufacturers shifted one product or process to another depending on the market needs.

Amin and Thrift (1995) have also described this period as the post-Fordist or post-industrial era, which is characterised by among other things, the diffusion of technology, volatile markets, and the growth of a large number of small- and medium-sized firms in parts of Europe. As the unstable market conditions curtailed returns on investments, e.g. machines often became redundant before the investment costs were recovered; many firms experimented with flexible working conditions and production methods. The
resulting high level of efficiency, which previously had been reserved for multinational large corporations that thrived on mass production, as well as active support from institutions, enhanced the growth of small- and medium-sized firms.

Following on this, “Marshallian districts” became the symbol and point of departure for understanding and analysing the dynamics of the “third Italy” (Cossentino 1996, Brusco 1986, Becattini 1995, Baker 1995, Brusco 1995, Dei Ottati 1994). This name is used to differentiate industrial districts in the south of Italy and the triangle of Genoa, Turin and Milan in the northeast and northwest of Italy from the rest of the country. These studies revealed the presence of many industrial districts that were home to small- and medium-sized firms with flexible production methods. The studies found that, for example, in the knitted goods industry in Capri, the ceramic tiles industry in Sassuolo, and the textile industry in Prato. Taken altogether, these districts numbered around 29 (Cossentino 1996).

In accordance with Marshallian dynamics i.e. specialised labour markets, learning etc., the Italian case studies showed that they had the capacity to attract skilled workers, which was one of the major advantages that Marshall (1920) had written about. Accordingly, the availability of jobs for people who want to take them and the level of entrepreneurship indicated that the districts were growth areas (see, for instance, Cossentino 1996, Brusco 1986, Becattini 1995, Baker 1995, Brusco 1995, Dei Ottati 1994). The division of labour between the Italian industrial districts allowed each district to have a unique form of specialisation. Workers and technicians move between the different firms, for example between suppliers and buyers, but they tend to stay within the district. In this way, tacit knowledge among the workers could be accumulated (Berggren et al, 1996, Becattini 1995, Brusco 1995).

In Sweden, the Gnosjö region located in the province of Småland in the southern part of the country showed the same pattern in terms of generating small enterprises. The small companies in this region were argued to combine competition and collaboration. Just like the Third Italy, the presence of socio-economic networks that created a breeding ground for entrepreneurship was observed (Wigren 2003, Brulin 2000, Johanisson 1984).

The ability of Gnosjö region to foster entrepreneurial activities and nurture a climate of trust between the economic actors through embedded relations that span through the community is referred to as “the Gnosjö spirit”. According to Gummesson (1997), the foundations of this spirit started in the sixteenth

40 The Marshallian district concept has also been used to analyse dynamics regions in many European, Asian as well as North America. In addition, in the 1980s and 1990s it was the dominating concept in regional development.

41 In Swedish “Gnosjöandan”.

116
century when a weapon factory was established in the scarcely populated, small-farming area. The iron industry and trade rose during the sixteenth and seventeenth centuries when foreign smiths taught new methods. After the death of the Swedish warrior king Charles XII and Sweden’s defeat in the Great Northern War, local blacksmiths witnessed a sharp fall in the demand for their metal products. Following the decrease in demand, many of the now self-employed blacksmiths started to expand their production lines. However, from the mid-nineteenth century, the wood industry and glassworks took the place of the iron industry.

In summary, the characteristics of the industrial districts of the Third Italy and the Gnosjö region\textsuperscript{42} can be summarised as follows (Baker 1995, Becattini 1995, Brusco 1995, Asheim 2000):

- Geographic concentrations of many small- and medium-sized firms.
- Dense socio-economic networks that encourage inter-firm relationships and enhance competitiveness.
- Institutional arrangements that facilitate the flow of ideas, which enhances development of production methods and market innovations.
- Willingness to adapt to new circumstances and flexible production systems.
- Division of labour between the districts allowing for high levels of competence and accumulation of tacit knowledge.

Meanwhile across the Atlantic, territorial competitiveness and the significance of the region as an important unit of development gained further importance when Anna Lee Saxenian published “Regional Advantage: Culture and Competition in Silicon Valley and Route 128” (1994). In this book based on a study of the two areas of route 128 in Massachusetts and Silicon Valley in California, she concluded that the success of the latter depended on the dynamic horizontal networks of managers, companies and supporting public institutions. In contrast to the static vertical integration of inter-firm relationships in Route 128, she found that the dynamic networks that existed in Silicon Valley provided a receptive atmosphere, which encouraged cooperation while still maintaining a degree of competition.

At the same time, several social scientists (For instance, Staber 1998, Markusen 1996, Mascanzone et al 2000, Baker 2000) began to develop

\textsuperscript{42} It is worth noting that although the Third Italy and the Gnosjö region are similar in the spirit of encouraging entrepreneurship and the many informal networks, there are also differences between the regions. For example, in the Gnosjö region, the industrial specialization is horizontal, while in the Third Italy it is vertical.
concepts that were largely based on the work of Marshall to explain spatial competition. They examined key features of special areas in terms of economic development and began to categorise the different kinds of industrial districts.

Staber (1998) argues that there are two distinct kinds of industrial districts. First, there are the “Marshallian” type industrial districts with competitive inter-firm relationships. These districts have, of course, firms that come and go because of forces of competition. New firms are born out of older firms, new entrants emerge as a result of needs within the districts, and firms are closed as a result of increased costs, lack of customers, etc. Secondly, there are industrial districts, mostly like the ones in the Third Italy, which have inter-firm relationships and are socially integrated. The dynamics arise as a result of institutional arrangements for collective learning and social relationships. Staber (1998) asserts that the backbone of industrial districts is the institutional setting with common, but varying, mechanisms of integration based on kinship, religion, and relationships. Actors are involved in dense social networks that facilitate not only competition due to peer pressure, but also co-operation through drawing on network resources.

In the USA, Markusen (1996) found the presence of industrial districts that were quite different in characteristics from those in Europe. In *Sticky Places in Slippery Space: A Typology of Industrial Districts*, she gives an in-depth analysis of neo-Marshallian industrial districts. Markusen (1996) contends that the research findings from studies of the Italian districts, which are based on the socio-economic networks of small- and medium-sized enterprises, are far from representative for other industrial districts throughout the globe. To capture differences in terms of networking, regional development and institutional context, she proposes two alternative models of industrial districts.

The first kind of industrial district is similar to Péroux’s growth poles where a dominant firm pushes the region forward, which is an important element in growth pole theory. Markusen (1996) calls it the *hub-and-spoke*. These kind of industrial districts consist of one or several major employers with supplier networks and service providers. These are often industrial areas that house single enterprises like the Boeing complex in Seattle, or larger areas that house firms in the same industry like the motor industry in Detroit.

The dynamics of hub-and-spoke industrial districts is centred on the ability of small- and medium-sized firms (SMEs) to interact in networks, and the capacity of a large firm (an engine enterprise) to accommodate the smaller firms. The SMEs are linked through transactional relationships to the engine enterprise, to which they sell to or learn from (Barkley 2001). The co-operation and competition between the companies are usually dictated by the engine enterprise. The small and medium sized companies in the hub-and-spoke gain external economies from each other by virtue of their number. For
example, at the Boeing complex in Seattle, several SMEs supply Boeing with various inputs (Dunning 2000, Markusen 1996). Mascanzoni et al. (2000) report the presence of sixty hub-and-spoke districts in the Third Italy. These are characterized by vertical and horizontal networking between the firms, which are mainly small suppliers and an engine enterprise. Small local networks of firms import, store and distribute raw materials necessary for the engine enterprise to produce finished goods. These drive the pace of product innovations in the district. These locomotive companies in the Italian sense are global players with international reputations and they are characterised by their excellent professional knowledge of craftsmanship, technical and social know-how, and innovation ability. Their success is a result of technological, organizational and marketing abilities that have been accumulated during many years of conducting business activities in the districts. Close proximity between the engine enterprise and their suppliers and customers facilitates the transfer of strategic information, which small- and medium-sized suppliers are dependent upon.

There is a two-way linkage of relationship between the firms in a hub-and-spoke district. First, there is a vertical relationship between the engine enterprise and the SMEs that are clustered, and, secondly, a horizontal relationship between the SMEs. Competition is limited because most of the SMEs supply different services or products, or they are all involved in the value chain of the engine enterprise (Barkley 2001).

The second category according to Markusen (1996) is the satellite platform and state anchored industrial districts, which often are induced by policy measures to enhance regional development. Defence installations, universities and concentrations of government agencies belong to this category. Policy measures play a vital role not only in their evolution, but also in their successful development. In these kind of industrial districts, there is no collaboration between the companies as each and every one of the tenants is engaged in different kinds of activities. Most of them are large players that determine the pace of development.

Although Marshall (1920) discussed the knowledge aspect of economic concentration, most of the studies on “Marshallian districts” from the eighties onward had their analysis in the institutional setting with a focus on cluster dynamics in explaining their evolution and functioning. The case studies of the Italian districts and other areas around the globe described the relationship dynamics, including the collaborative arrangements between economic actors and public bodies in achieving development. This aspect of cluster based relationships as a facilitator of regional development was to a certain extent discarded by mainstream economists.
Post globalization concepts

In the early 1990s, spatial competition models including Marshall’s were aligned with new growth theories, mainly due to the importance of knowledge which came to be viewed as a factor of competition. These theories hold that steady economic growth in a nation or a region depends on technological progress and knowledge utilisation (Glaeser 2000, Ekstedt 2001). Accordingly, the development of new technologies and their diffusion by way of knowledge accumulation are critical components in a company’s productivity and henceforth on the context. Much of the literature builds on the works of Romer (1990), in which he models the micro-economic environment from a knowledge accumulation perspective and the underlying theme is that the development of the individual firm depends on knowledge accumulation and, consequently, its productivity and innovative capabilities depend on the extent of knowledge form creation and proliferation.

At the same time, the new growth theories brought forth the impact of research, development, and the spin-off effects of knowledge-based firms in the context of location. Accordingly, knowledge accrual is often the result of education, experience and training that is contextually generated and diffused. Therefore, much of this research stream attempts to explore the geographic dimension of economic development by looking at mechanisms that lie outside the borders of the particular firm, including the extent of knowledge spillovers between actors. These are mainly the mechanisms that were earlier discussed by Marshall (Laestadius 1999).

The recognition that knowledge creation and diffusion is an important mechanism to achieve economic competition and territorial development coincided with a period that was marked by political and economic changes. The collapse of the Berlin Wall and the subsequent opening up of new markets in Eastern Europe and the expansion of the European community (later on, the European Union) are just a few examples of the political developments. On the economic front, globalisation, improvements in transportation and information flows mainly in the telecommunications industry and the disintegration of the larger manufacturing industry in the developed world (Castells 1999) led some to predict the “death of distance” (see Cairncross 1997). While others, e.g. Amin and Thrift (1995, p.2), argued that globalisation” does not represent the end of territorial distinctions and distinctiveness, but an added set of influences on local economic identities and development capabilities”.

The developments mentioned above gave an added significance to knowledge as a catalyst of economic development and heralded the evolution of new

---

43 This is sometimes dubbed as endogenous growth models.
concepts such as “the knowledge economy” (Castells 1999), “the learning economy” (Lundvall 1992) to capture the knowledge aspect of development. The spatial dimensions of these changes were captured by concepts such as the advent of a “new economic geography”. In the early 1990s, Paul Krugman for instance developed his trade theory of “the new economic geography”, in which he constructed an equilibrium-based model of trade to explain the concentration of business activities in a given place. His model of interplaying factors contained: increased returns to scale, transport costs and factor mobility. According to this model, increased return to scale motivates producers to limit production costs, transport costs are reduced because of the geographic proximity, and factor mobility is achieved because of the expansion of the labour market (Krugman 1996). Tson Söderström, et al, (2001) in a much publicised report in Sweden, for example employ the metaphors of “forces of concentration and forces of dispersion” to describe the features of the new economic geography. On the one hand, the new economic geography is manifested in the replacement of traditional factors such as economies of scale by relationship-based factors (forces of concentrations). On the other hand, according to Söderström et al, (2001) the new economic geography is characterised by the forces of dispersion. These include the presence of infrastructure, which attracts skilled labour, research and development, customers and suppliers.

In addition, since the “new” concepts reveal a positive relationship between innovation, knowledge and economic development (Lundvall 2001, Archiburgi et al 2001, Storper 2000a, Audretsch 2000a), they had an influence on industrial policy. Thus they contributed to a rethinking of regional and national development strategies, especially at a time when previous policy measures appeared to be unsustainable. The relative successes of industrial districts and the digital boom orchestrated by the advent of the Internet have also certainly enhanced the development of new frameworks of competition; a paradigm that is by and large based conceptual-based development (Raines 2001). In the period leading up to the new millennium, there was a widespread debate on the “knowledge economy” as a mechanism that would contribute to regional and national development. In the report A New Economy? (2000), the OECD described knowledge creation as the ultimate panacea to overcome the challenges of the global economy. In terms of regional development, the debate stressed the importance of codified forms of knowledge that could be accessed and generated through institutions of higher education.

The new regional development policy, which evolved from these discussions is referred to as endogenous regional policy, which simplified means a departure from the Keynesian inspired convergence model aimed at achieving regional parity. There was also an emphasis on a systems-based strategy to enhance regional competitiveness (see, Nuur et al, 2009 and Nuur and
Laestadius, 2010). By adopting a regional/national innovation systems’ approach the government hoped Swedish industry would become internationally competitive. The theoretical undercurrents of the concept of, “National Innovation Systems” (NIS) related to the work of Christopher Freeman (1988) who described the success of Japan. Freeman (1988) discusses how industry and public policy interplayed to contribute to innovative methods, process and products. The NIS concept has been developed by, among others, the economist Bengt Åke Lundvall. Central to the notion of NIS is the idea that differences between nations in terms of economic development depend on a myriad of factors including the degree of knowledge utilisation, institutions and infrastructure rather than on basic factor endowment:

“National innovation systems are, by definition, localized and immobile, and thus able to provide firms with valuable capabilities and framework conditions not available to competitors located abroad, even under the most open market conditions imaginable.” (Lundvall 2000, p.364).

The relative success of Silicon Valley and Route 128 in Massachusetts in fostering entrepreneurship and knowledge-based dynamics (Saxenian 1994) and the political as well as economic changes mentioned earlier increased the importance of innovation systems as a developmental tool. In Sweden, the central government has with the establishment of Vinnova (the Swedish Agency for innovation systems) embraced an industrial policy with the goal of enhancing the competitiveness of Sweden as a nation by promoting the innovative capabilities of industry. Following this, the parameters of the national innovations systems approach were expanded to include actors at the regional level with the advent of regional innovation systems (RIS) (Bager-Sjögren & Rosenberg 2004).

In Sweden, the seeds of the regional innovation systems approach which are discussed by Rickne and Laestadius (2016 – in this volume) were sown in the 1970’s when higher education was decentralised (Andersson et al, 2004). The establishment of universities and university colleges in almost every political province (län) and the development of science and technology parks in the 1990s is a testimony of this approach. The underlying implication of this policy approach is a belief that knowledge spillovers between geographically proximate actors or functionally interacting actors would have a positive effect on regional development and contribute to the emergence of knowledge-based firms that would replace the decline of traditional sectors.

**Industrial Clusters**

If the national and regional innovation systems approach is one that answers the “how’s” of knowledge generation, one concept that has its point of
departure in the “whereas” of development is “Clusters”. In the 1990s and the earlier part of 2000, this concept had become part and parcel of the regional and national economic development debate (see, for example, Stymme 2004). In the economic, geographic and business literature, the phenomenon of cluster was too often used interchangeably with industrial districts and science and technology parks. In the past, scholars have used the perspective of geographic agglomeration, which Marshall pioneered to describe clustering of economic activities. This description denotes the tendency of firms in the same field to locate in the same geographical area. In most circumstances, however, the physical location of firm agglomeration is based on factor endowment and does not necessarily mean the presence of interdependencies (Malmberg 1998).

**Cluster dynamics**

As indicated earlier, clustering of economic activities is not a new phenomenon, but has rather occurred for centuries. However, it gained potency in 1990 when Michael E. Porter, Professor of Business at Harvard University, authored a book entitled “The Competitive Advantages of Nations”. The findings in this book are based on research carried out in several European countries, including Sweden. The Swedish analysis is contained in a book entitled “Advantage Sweden”, which was co-authored by Örjan Sölvell and Ivo Zander at the Stockholm School of Economics. For the purpose of this chapter, I intend to embrace Porter’s (1990) definition of a cluster, which is:

> Geographically proximate groups of interconnected companies and associated institutions in a particular field, linked by commonalities and complementarities. The geographic scope of a cluster can range from a single city or state or a country or even a group of neighbouring countries.” (Porter 1990, p.254)

The origins of clusters are the same as those of agglomerations and include historical circumstances, proximity to sources of endowment, local demand conditions and random elements (Porter 1990). However, not many of these factors act independently, but rather they interact with one another. Technology is one such factor. Krugman (1996) argues that technology is a vital factor that induces clustering, but not the main one. Instead, technology helps the pace of the cluster development, as it takes some time to discover and diffuse it. Technology has been cited as one of the factors that contributed to the growth of the well-known cluster of Silicon Valley. The discovery and subsequent diffusion of the semiconductor and the computer processor certainly played a leading role in the development of Silicon Valley (Saxenian 1994).

---

44 See for example the discussion earlier Marshallian districts.
Tson Söderström et al (2001) assert that clusters go through various stages that range from “the heroic” to the “renaissance” phase. The heroic phase is the initial period of clustering and is characterised by the presence of a few actors, mainly entrepreneurs or a flagship enterprise. During this phase, there is almost no vertical relationship between the actors. The second phase of clustering is marked by the presence of dense networks of relationships that are vertically integrated and a critical mass of firms that are concentrated in the same area. The final phase is characterised by “diversity” and specialisation, with many sub-clusters and firms that are vertically integrated.

Rosenfeld (1997) also argues that economic clusters have a life cycle that is marked by embryonic, growth, maturity and decay stages. The embryonic stage of a cluster is the result of technological innovations and/or inventions and internal direct investments made either by an individual or a firm. The growth stage is marked by the developments of markets, the spin-offs of firms, the attraction of imitators that develop unique products, competition that allows for benchmarking and the facilities to attract entrepreneurship. The mature stage is marked by new entrants, including service providers, shared infrastructure that leads to cost efficiency for the companies and the ability to combine competition and cooperation. The decay stage takes place when the products that the cluster produced are replaced by cheaper production elsewhere or substituted for as a result of new innovations.

When introduced in the early 1990’s by Porter, the cluster approach defied traditional models of the location of economic activities, which were mainly based on comparative advantage, and were static in their nature (Malmberg et al 1996). In contrast to agglomeration theories which are mainly embraced by economic geographers, Porter (1990) took a business science perspective and contended that it is the interplay between geography, institutional linkages and inputs from formal and informal organisations that forms the basis for competitive advantage of nations. In addition, this perspective recognised the importance of interdependencies between firms in terms of relationships and knowledge building in developing production competencies (Cooke 2002). Porter never described how large a geographical area ought to be, or how many firms ought to be concentrated in one specific area to be called a cluster (Berggren & Laestadius 2003). However, one vital element of his thesis was that firms do not live in isolation, but rather depend on other firms and organisations for various kinds of resources that enhance productivity, innovation and new business formations.
The diamond model

Having previously written about firm-based competitive strategies, including value-chain management and firm-level competitive strategy through cost reductions, Porter (1990) developed a diamond-shaped model to describe the main elements that contribute to the competitiveness of nations. The model consists of four major interplaying factors that determine territorial competition: demand conditions, factor conditions, firm strategy, and related and supporting industries. It also included two exogenous variables that could be accessed by any nation or region: chance and government policy. Two of the components of the model, demand and factor conditions, are external mechanisms. These lie outside the borders of the firm and relate to the availability of a demand market and the production factors necessary to meet the demand. Accordingly, the basis for competitive strength of a nation is the presence of factor conditions, such as provision of upgradeable natural resources, capital and labour, combined with market demand conditions, such as local and national demands.

Porter (1990) relates the demand and factor conditions of nations to two micro-economic-based elements: individual firm strategies and the presence of related, and supporting industries. These two elements are firm- and branch-specific factors and form the basis of the cluster concept. The idea behind this is simple; Rival firms and other firms in the same field tend to appear in a cluster where they have access to common services which reduces transactions costs and allow for the provision of human resources. The presence of related and supporting industries facilitates a process of innovation and value-creating mechanisms, without the individual firm needing to make all the investments on its own. The interdependencies that arise in terms of inputs, such as training and specialised labour, are managed collectively. In this way, the firms can significantly reduce their production costs as compared to rivals that are located in different places.

An analytical instrument and a policy tool

The concept of economic cluster gives an analytical opportunity to scholars, as well as to policy makers, to capture the mechanisms that lead to industrial development in a nation or a region (Malmberg 2002; Brown 2000). Through the dynamic linkages of relationships and transactional activities, the argument went that the cluster concept helps to identify competitive sectors of the economy (Malmberg 2002). In addition, it provides scholars with instruments to study how firms perform in an institutionalised system rather than seeing the firms as static units usually described by statistic variables of

45 For an in-depth analysis of this model on Swedish industrial dynamics, see Advantage Sweden (Sölvell et al, p. 24)
size, number of employees, etc. According to Malmberg (2002), using the cluster approach provides insights into functional linkages such as:

- Vertical transactional links, i.e., exchange of goods, know-how, etc. between firms, customers and suppliers.
- Horizontal competitive links, i.e., competition between firms in acquiring factors of production such as labour.
- Knowledge spillovers through interaction between actors.
- Horizontal co-operation through formal and informal contacts such as strategic alliances, networking, etc.
- Commonalities such as technologies, complementary activities, infrastructure and labour pools.

As illustrated by Malmberg (2002) above, using the cluster approach to analyse regional or national development poses measurement challenges since functional linkages are subjective and difficult to quantify. According to Berggren and Laestadius (2003), the adoption of a quantitative cluster approach to study industrial development is a complex one, since the extent of relationship connections between proximate economic actors is hardly accessible in industrial statistics. Instead, they propose a qualitative evaluation, which considers horizontal and backward linkages in the form of “development pairs”. Berggren and Laestadius (2003) exemplify this with the Finnish-Swedish cluster of the telecom companies Nokia and Ericsson, which they argue developed parallel with a distinctive feature of competitive rivalry.

In addition to being an academic analytical instrument, the cluster concept had also become a popular tool among policy makers to enhance regional and national development (Sadler 2004; Gordon & McCann 2000, Raines 2001, Stymme 2004). From a policy perspective, an important factor of the concept is the degree of co-operation that is needed to manage interdependencies between public institutions and organisations and more importantly policy measures to provide auxiliary inputs to companies to raise their competitiveness. As Waits (2000) contends, adopting a cluster-based approach allows policy makers to have insights into the factors that drive the economies of particular regions and the challenges that face them.

O’Mally and Van Egeraat (2000) examined how a cluster-based approach helps us understand regional and industrial development in Ireland. Despite an active policy strategy to induce and promote the growth of new industry using a cluster-based approach, they found that the performance of Irish traditional firms (Food processing, manufacturing, service, etc.) was not weaker than the performance of the designed clusters à la Porter.

Raines (2000) puts forward insights into how the cluster concept is embraced as a developmental approach in seven regions across the European Union (Art
Vally in France, East Sweden, Limburg in the Netherlands, North Rhine Westphalia in Germany, Pais Vasco in Spain, Scotland and Tampere in Finland. In general, he found that a cluster-based approach is influencing regional and national development policy either by complementing existing development policies or through using it as a new framework to achieve competitiveness.

Brown (2000) examined the literature on clusters and cluster policy and argues that there is a lack of coherency in both how the concept is defined scholarly and how policy makers apply it in practice. While some European countries such as the Netherlands have adopted a “top-down” approach to implementing clusters, other countries including Scandinavia have opted to have a bottom up strategy where cluster development is a local/regional issue.

Brandt (2001) has also studied cluster-based approaches in Sweden and the other Scandinavian countries and he contends that Porter’s cluster approach has shaped industrial dynamics in Sweden. He identified two basic measures to develop cluster dynamics: top-down and bottom-up approaches. These two approaches are based on providing direct and indirect support to devising cluster-based strategies. The direct approach is marked by a strategy of “picking winners” and “implicitly” targeting available competitive sectors. The indirect approach strategy involves devising programmes, such as “partnership” programmes, and endogenous development initiatives to induce cluster dynamics.

Despite the popularity the cluster concept has gained in both policy and academic circles, it is laden with ambiguities in terms of scope, contents and definitions (Malmberg 2002). For instance, in an article entitled “Deconstructing Clusters: Chaotic Concept or Policy Panacea?” Martin and Sunley (2003) discuss how “Porterian” clusters are shaping economic development policies. The authors assert that the concept is “vague” as to both definition and scope and that it contributes to “conceptual and empirical confusion” (2003, p.10). According to Martin and Sunley (2003), much of the cluster debate centres on policy discourse to find “fit” strategies to enhance development and that from a scholarly perspective it does not inject new blood into location as a matrix of competition. Their critique is based on how the concept is defined (they cite 10 different definitions) and that the political discourse surrounding it, as a tool to encourage innovation, clouds the scholarly value and the lack/difficulty of empirical foundations (Sadler 2004). In respect to the policy dimension, Martin et al (2003) suggest the role of

---

46 In Swedish: Östergötland.
policy should be to encourage productivity in all firms and caution against a policy that is committed to a “mind-set” of creating clusters.

Also, Feser (1998) points to the inconsistency of cluster definitions and the lack of clarity in understanding the dynamics that lead to the development of clusters. He argues that this results in an inconsistency in policy measures on cluster developmental approaches.

In a study commissioned by ITPS, the Swedish Institute for Growth Policy Studies, Malmberg (2002) follows this path of criticism to a certain extent. He asserts that it is unclear whether the term denotes the functional dynamics of geographically proximate organisations or the geographical scope. He argues that the ambiguity of the concept is strengthened by the policy discourse designed to enhance regional development. In policy terms, a cluster is “formed when an actor (often with a base in a public organisation, or private supporting institution) identifies a cluster, existing or potential, gives it a name (often with the suffix of “valley”) and starts to act to strengthen, develop and market it” (Malmberg 2002, p.16).

Apart from the companies which form the core, a cluster also encompasses public or non-governmental organisations that provide inputs such as competence development, education facilities, research and development, marketing and other business development mechanisms that make the cluster competitive (Porter 1998, 2000). However, although public policy towards encouraging cluster development is vital, according to Enright (2000) the role of policy is limited to improving rather than initiating and implementing. Apart from providing public goods and competence generation mechanisms, active policy participation aimed at “creating” a cluster could impede cluster development, as noted by Enright and Ffwocs-Williams (2000, p.4):

> “Guidelines on policy towards clusters must be premised on a view of government as supporting existing and emerging clusters rather than trying to create them from scratch. A policy aimed at developing entirely new groups of firms in selected sectors can entail high costs, high risks, serve as a screen for outmoded forms of industrial targeting.”

McDonald & Vertova (2001) also maintain that policy arrangements are beneficial in identifying potential clusters, to facilitate co-operation and competition, provide linkages between firms and support institutions as well as develop infrastructure. However, active policy participation in creating clusters from scratch and imitations could restrain their development and lead to lock-in effects.

In Sweden, in a much-publicised report entitled “kluster.se”, Tson Söderström et al (2000) caution against the proliferation of strategies that aim at creating
clusters in all regions of Sweden. While the authors maintain that Sweden has the capability to generate dynamic clusters outside the Greater Stockholm area”, they caution against “spreading resources too thinly over Sweden” (2000, p.23).

On the other hand, Rosenfeld (1995) takes a policy perspective on the development of clusters and contends that there is a need for active policy intervention to promote cluster-based strategies to achieve economic development. He argues that public policy should not only focus on building the necessary infrastructure, but also on “building” competitiveness and innovation by, for example, investing in social capital and learning. Furthermore, he asserts that cluster definitions have been complicated by the political discourse about networking and other inter-firm dynamics.

**Some classifications of clusters**

As indicated earlier in this section, Porter (1990) did not explicitly discuss the geographical scope of a cluster. This creates a size problem, especially since the concept has become embedded in the industrial development debate. The business economist Michael Enright (1996), however, is one of several researchers who have tried to describe the geographical dimension of clusters. He has suggested the following four varieties based on the works of Porter (1990):

- **Industrial clusters.** In this case he uses Porter’s definition of a concentration of companies in the same field and proximate institutions. This implies that basically the geographical scope of this cluster could be more or less anything. The reason why Enright uses the word “industrial” in this respect can be understood given that Porter’s studies are based on nations rather than regions or small geographical areas.

- **Regional clusters.** Concentration of firms in the same or related branch and regionally based proximate institutions. Paradoxically, regional clusters allow firms to have a measure of both co-operation and competition, as they are outward-oriented, such as the firms in the Gnosjö region. Isaksen (2001, p 104) argues that the term “regional clusters” encompasses a broad description of “industrial districts, innovative milieus, local industrial complexes and new industrial spaces”

48 There are several ways to classify clusters. Different researchers have provided different typologies of clusters and my ambition with this study is not to give a thorough typology. Instead this classification merely attempts to shed some light on the topic and is not meant to offer a solid typology.
• Industrial districts. Firms that are geographically bound and engaged in complementary and common production processes, such as the firms in the Third Italy.
• Business networks. Firms that are not necessarily located in the same vicinity, but that have continuous communication and interaction with one another. Although the actors have relational bonds and conduct transactions with one another, this kind of cluster is functional rather than geographical, and it does not necessarily mean that they are in the same or related branch.

Park (2000) identified six kinds of clusters based on scope, technology and division of labour. These are “three basic types of Marshallian districts, hub-and-spoke, mature satellite, and pioneering high-tech industrial districts” (Park 2000:329). Dunning (2000) uses a classification also with six varieties, similar to that of Park:

• One category is the hub-and-spoke cluster, which is induced by the presence of an engine enterprise. The dynamics of this cluster evolve around gaining external economies through supplier and customer networks.
• Another category is industrial districts, such as the districts of the Third Italy. This cluster is characterised by factors of production, such as access to labour and natural resources. The dynamics of the industrial districts are achieved through external economies in terms of reduction of transportation, transaction and labour costs. Three of the categories are clusters induced by policy measures to create learning and innovation opportunities. These are:
  • Institutional clusters that are formed as a result of their proximity to institutions of higher education and evolve around universities.
  • Public bodies that provide public infrastructure. Examples are Farnborough and Aldershot in the United Kingdom, which are centres for aerospace and atmospheric research.
  • Generic clusters that mostly are started as a science or technology park. Their success depends on sophistication, upgrading of infrastructure and an innovative milieu that encourages growth through knowledge fluidity and learning externalities.
  • The last category is export zones, which developing countries have established in a bid to improve competitive strengths and to obtain foreign currency. Multinational companies are invited to make a presence in these clusters.

Rosenfeld (1997) takes a relationship perspective and identified three kinds of clusters based on relationship densities.
• **Working clusters.** This is a mature agglomeration of firms that recognise their interdependencies. The firms and the supporting institutions of the cluster are involved in interactive processes of learning and innovation. The social structure allows for exchange of information and ideas, business start-ups and networking, and it seeks to strengthen the systematic and economic value and image of the home region. This type of cluster, which is sometimes called “overachieving” clusters, is found in the Silicon Valley, the ceramic tiles industry in Italy, and the horticulture industry in Holland.

• **Latent clusters.** This kind of cluster is referred to as “underachieving” clusters. These clusters have the potential to develop and become a working cluster, but so far there is no social structure that allows for a free flow of information. Despite the presence of firms and supporting institutions, interaction between actors is weak. Although the actors in the clusters are sharing much, they are not aware of the interdependencies that exist between them. They also lack supporting organisations that actively seek to create an arena for networking and interaction. One such cluster is the biotech cluster in the research triangle in North Carolina.

• **Potential clusters.** This kind of cluster is also known as “wannabees”. They have many characteristics of a cluster, but lack competitive advantages. These clusters are often found in the technology sector and they are often initiated, designed and selected by policy makers. Another type of potential cluster is an agglomeration of a group of firms in the same or related industry that lack political power, skilled workers or exposure. Examples of such clusters are the rural cluster of wood products in eastern Oklahoma and microbreweries in Oregon.

**Creativity and location**

In the last few decades, the notion of creativity has emerged as concept in industrial location and in the policy regarding regional and national development. If Michael E. Porter was the sage of competitive advantage by devising the concept of clusters and its dynamics at the beginning of the 1990s, Richard Florida has given potency to the old notion that creativity is the precursor of competition and in – locating creativity to large urban areas – also become “the cool-cities guru” of the new millennium (see Peck, 2005 for an overview).

---

The point of departure of Florida’s creative class thesis is that the global economic integration witnessed in the last decades makes obsolete the conventional wisdom that differences between nations, regions and cities depend on classical comparative advantages. Competitiveness – and here, we are close to the Porter (1990; 1998) position – is a social and cultural construction. Florida focuses on creativity, which according to him differs and develops differently between cities that are assumed to be its loci. For example, in the “Rise of the creative class: And How It’s Transforming Work, Leisure, Community and Everyday Life” (2002c) he argues that in the era of globalisation US cities with diversified economic activities and diversified human capital formation have experienced the advent of a creative class while others have witnessed the erosion of creativity. The creative class uses technology as a channel to produce/sell products, penetrate markets and communicate (Florida, 200b).

The members of the creative class are assumed to be attracted to milieus that provide them with amenities that nurture their creativity and conform to the phases in their life projects (Florida, 2004). Their mobility is thus conditional and influenced by the erosion of the demarcation lines between work and leisure. The creative class, according to Florida, constitutes a significant part of the population of cities that are characterised by diversity and tolerance in terms of ethnic, cultural, religious and sexual orientation (Florida, 2002c).

In short, the creativity thesis put forward by Florida (2002a; 2002b; 2002c; 2004 and 2005) could be summarised as follows:

- Competitiveness depends on success in recruiting talent – and this is true more than ever in the knowledge based economy.
- There is a new and intensive global competition on talent – due to globalization as well as to the rise of new Tiger economies.
- The highest probability to get into contact with these talented workers is in highly urbanised regions.
- Talented people have strong expectations based on lifestyles, which have a significant impact on their willingness to follow job opportunities; they optimize their whole life project; closeness to the opera may be as important as career options. Consequently, firms have to follow their staff to the triple T –regions (Technology, Talent and Tolerance), if they want the best ones, rather than the other way around.
- Triple T regions may develop a certain self-propelling dynamic incorporating a lifestyle oriented economy (galleries, theatres and meeting points) highly integrated with the lifestyles of the creative class.

Thus – following the above- it could be concluded that regions and cities that cater to the needs of the creative class win while those that do not face a melancholy demise. Overall, to policy makers in industrialised nations the
notion of creativity as an instrument of development is appealing since it is not necessarily footed in certain regions, nations or industries. In addition – superficially at least – it opens the way for creativity enhancing policies primarily in areas in decline. Is this analysis reasonably plausible and is the process unidirectional, or is there a potential variety as regards the development paths towards the knowledge-based economy? Will non-urbanised regions in the old industrialized countries die while larger urbanised ones prosper? Is the role and behaviour of the creative class a general or universal phenomenon independent of time and location or is this a process related to specific historical or geographic conditions or relevant for certain life cycles of industries or technologies? And more importantly what do we mean by “creativity” and “the creative class”?

**Creativity- a Contested Concept?**

The characterisation of Florida´s creative class is loosely based on occupations (engineers, architects, Scientists, economists etc.) (Florida, 2002c, 2004). It also contains “bohemian” individuals who pursue lifestyles that add to the variety of elements in the creative class concept (Florida, 2002a). In the end, however, it might be the case that the creative class is both wider than the set of citizens that have a Bachelor’s or Master’s degree (many bohemians do not) and more narrow: teachers, social workers and pilots are e.g. not included in the creative class (see e.g. Markusen, 2006 for a critical review of this). However, close to a study by Janik and Toulmin (1973) on the creativity of Wittgenstein’s Vienna (although not referring to it) Florida (2002a) has examined how the geography of bohemians corresponded to the geography of diversity (immigrants, gays), talent (defined as those with at least a Bachelor’s degree) and high-tech industry location. This conclusion has been questioned by other scholars, however (cf. e.g. Markusen, 2006 & Wojan et al, 2007).

In a wider sense, creativity as a concept is not only solely based on the level of education attained by individuals. It concerns idea generation (cf, Franken, 1994; Stenberg and Lubart, 1999; Clark and James, 1999; Boden, 2004). It is, as defined by Robert E. Franken, (1994, p396) “the tendency to generate or recognize ideas, alternatives, or possibilities that may be useful in solving problems, communicating with others, and entertaining ourselves and others” or as Boden (2004, p1) puts it “the ability to come up with ideas or artefacts that are new, surprising and valuable”. “Ideas” here include concepts, poems, musical, compositions, scientific theories, cookery recipes, choreography, jokes- and so on. “Artefacts” include paintings, sculptures, steam engines, vacuum cleaners, pottery, origami, penny whistles – and many other things you can name.

From the above approaches, it is clear that the notion of creativity relates to the generation and appropriation of novel ideas by individuals (Franken, 1994; Stenberg and Lubart, 1999; Boden, 2004). Hence, it is not characterised
by the level of education of the individual but by a process of spontaneity in
its inception and sustainable outcomes although Clark and James (1999)
assert that it can also be oriented endeavour and as such depends on previous
knowledge. Even at this stage, Rehn and De Cock (2007) argue that the
literature on creativity focuses on how it evolves without paying much
attention to the potentials. Creativity may be limited to certain areas of
knowledge as well as spilloverspillover between such areas. It is, however, far
from easy to show that creativity within certain aspects of the arts has any
relation to creativity within e.g. high-tech industries (cf. Markusen, 2006 and
Wojan et al, 2007).

Limiting our perspectives to only Florida’s approach might lead to a failure to
counter the importance of context/community related tacit knowledge and
the social capital that underlies much of the generation and dissemination of
creative ideas that is relevant for industrial transformation. Industrial
creativity does not happen in solitude but instead takes place through
interplays with the result of cumulative learned actions and existing industrial
practices and technical solutions developed among people that do not
necessarily have academic degrees. Knowledge creation and diffusion often
takes place through vast formal and informal networks that interact and
provide ancillary inputs which contribute to nurturing existing capabilities
and the fostering of new ideas. This entrepreneurial creativity was formulated
already a century ago by Marshall in the famous line “(...) If one man starts a
new idea, it is taken up by others and combined with suggestions of their own;
and thus it becomes the source of further new ideas” (Marshall 1890/1920,
p.271). If this kind of – maybe traditional or old fashioned – creativity is losing
its importance then phenomena related to such a historical transformation
should be included in the analysis, not simply neglected. From the context of
regional development, creativity should be put in the social context where it is
induced, generated and diffused. As such the term involves the capacity of
actors to individually or collectively breed ideas (novel or not) which
eventually promote competitiveness in the environment or context where they
are introduced.

Creativity as a concept – not least as formulated by Boden (2004; see above)
– has a clear connection with innovativeness. Readers familiar with
innovation theory may recall the Schumpeterian definition (1934/1968) of
innovation as creative combinations. Already for Schumpeter (1943) it was
clear that the “new” could be a surprising combination of something “old”.
And he used the term creative destruction to capture the renewal capabilities
of capitalistic societies (Schumpeter 1943). Inspired by Schumpeter, students
of industrial and technological transformation have since then viewed
creativity as an important pillar of the mechanisms that contribute to renewal
(e.g. Dosi et al., 2000, Nelson, 1994).
There is also, following Boden (2004) an obvious connection between creativity and the worlds of design and engineering. This has, among others already been formulated by Simon (1962, 1996) who relates these activities to the “sciences of the artificial”. What engineers (and designers) do is creating – through design processes – artificial worlds which are largely based on existing knowledge which is reorganized and synthesized into new forms. Putting aside, the conceptual discussion on the notion of creativity and limiting ourselves to spatial development, Florida’s perspective relates to the abilities of technology, talent and tolerance in creating the conditions for development. This is vital and worth analysing. While Florida’s “tolerance” might not have been on the agenda half a century ago when the economic thought governing regional differences was coined, technological diffusion making industrial production redundant and the erosion of talent in the form of brain drain were two issues that were even then seen as crucial in regional economic transformations (Myrdal, 1957, Hirschman 1958, Hansen 1966/1972, Lausen 1969, Clark and Whiteman 1983). Also discussed by Gunnar Törnqvist (1983) and Andersson (1981) the core theme of this literature was that an innovative environment facilitates knowledge transmission, competence development, regional learning and as such provides synergies to innovate product, processes and methods that could be commercialised. Evidently, Florida’s suggestion of the magnet effect of cities discussed in “The rise of the creative class: And How It’s Transforming Work, Leisure, Community and Everyday Life” and the repulsive effects discussed in “The flight of the creative class: The New Global Competition for Talent” although devised from the US context, could be applicable to regions and as such match Myrdal’s spread and backlash repercussions on peripheral regions. Thus, the rise of the creative class in relatively urbanised regions is matched by the fall of the creative class in peripheral regions since in absence of employment opportunities young people in particular tend to migrate to urban regions in search of better living conditions as stipulated by Myrdal (1957). At the same time, we are aware that Florida’s creativity notion centres on an age-old question of the mechanisms which promote and foster knowledge formation. Students of economic geography have for decades discussed whether it is regional specialisation or diversity that nurtures knowledge creation and diffusion. The specialisation argument takes its point of departure from the contributions of Marshall (1920), Arrow (1962) and Romer (1986). Known as the MAR perspective, the underlying premise is that economic growth depends on the extent of knowledge utilisation and technological progress. Scholars using this perspective model the micro-economic environment from a knowledge accumulation perspective and argue that firm competitiveness, productivity and innovative capabilities depend on the extent of knowledge transfer within organisations and knowledge spillover to other actors (e.g. Romer, 1986; Feldman and Audtretsch, 1999). It is important here to note that the MAR
supposition also captures the importance of location, as it reveals the economies associated with R&D and as such explores the geographic dimension of economic development by looking at mechanisms that lie outside the borders of the particular firm, including knowledge spillovers between actors.

The creativity thesis put forward by Florida as an element of development, on the other hand -manifested in the three T:s that make up Florida's thesis – Technology, Tolerance and Talent – relate strongly to Jane Jacob's (1969) study in which she argued that cities have magnetic abilities to attract people and industries that perform diverse operations i.e. the opposite of the MAR thesis. But it also has a family resemblance not only with the previously mentioned study by Janik and Toulmin (1973) but with a Swedish study entitled K-samhällets Framtid (Andersson and Strömquist 1988) in which it is argued that the four Ks; Kunskap (Knowledge), Konst (Art), Kreativitet (Creativity) and Kommunikationer (Communication) are the cornerstones of the knowledge society. Andersson and Strömquist (1988) argued the future growth of that society was based on developing webs, the hubs of which are found in the rapidly growing K-regions with strong similarities to Florida’s T-regions.

Jacobs (1969) argues that urbanised environments provide the nexus of economic development. In The Economy of Cities, Jacobs (1969, p1) poses the question “why some cities grow and why others decay”. In contrast to the suggestions of the MAR arguments, such as knowledge spillovers due to specialisation, Jacobs argues that the varieties in skills emanating from the creative urban context are the decisive elements. The Jacobs’ diversity supposition gained approval following several studies contrasting it with the MAR perspective; For instance, Glaeser et al., (1992) compared the MAR and Jacobs hypothesis by using data from 170 cities in the United States and found it consistent with the Jacobs position that it is the variety that cities offer and not the specialisation of labour/industry that explained knowledge spillovers. Feldman and Audtretsch (1999) also tested the diversity versus specialisation thesis by asking the question “does the specific type of economic activity undertaken within any particular region matter?” and found that industrial diversity is conductive to knowledge spillovers. Desrochers (2001) adopting a Schumpeterian perspective also found support for Jacobs’ externalities concerning the creative combinations of existing unrelated things promotes human creativity and innovations in regions. From a policy perspective Desrochers (2001) cautioned against policy strategies to promote the emergence of regional specialisations.

In the Swedish context, several studies placing the creativity thesis in a Swedish context have emerged; for instance; Andersson et al., (2005) confirmed Jacobs’ hypothesis with regard to commercial patenting in Sweden (1994-2001). According to this study, the propensity to engage in commercial
patenting activities was larger in urbanised regions that have diversified economic activities. Also Melander and Florida (2007) used a path analyses and measured traditional variables of human capital formation (wages, education level etc.), technology and talent and found that the three relatively urbanised cities of Stockholm, Gothenburg and Malmö accounted for 50 percent of the Swedish creative class. Stockholm scored highest with 30 percent, followed by Gothenburg 11.6 percent and Malmö 6 percent.

Tinagli et al., (2007) defining the creative class as the “the share of workforce engaged in conceptual and creative types of occupation, like managers, scientists, architects, engineers, artists, entrepreneurs, and many others” argue that overall there is a north-south divide in Sweden. While urbanised counties of Stockholm, Gothenburg and Malmö account for 60 percent of the Swedish creative class by scoring high in the tolerance (64 percent of the immigrant population are domiciled here).

**Conclusions and implications**

There is a myriad of literature on the underlying mechanisms of industrial location and the policy implications relating to that. This chapter has provided a brief overview of some of the key concepts. In summary, while the old literature sought to discuss location from an equilibrium approach, followed by the literature on the evolution of the centre as well as the periphery which put location in the context of regional development. In recent years, we have witnessed the emergence of concepts such as industrial districts, clusters, creative cities, etc. which view the economies of location and the policy implications could be summarised as embarking on a concept driven regional development. The underlying conjecture is that, there is a “path dependency” and territorial “lock-in” which means that history matters in understanding industrial dynamics and the development of an industry in a particular location. These relate to the notion that regions and locations that have similar historical backgrounds would experience similar patterns of economic development in the future (Malmberg & Maskell 1997). Accordingly, a territory’s ability to develop and compete with other territories is conditioned by history and comes because of cumulative decisions made by companies and policy makers.

Not least the onset of globalisation has brought a shift in regional development. While earlier, Keynes dictated the goals of regional development i.e. regional convergence with the goal to offset regional disparities, globalisation brought challenges and resulted in an endogenous approach. This approach is focused on regional growth and centres on inducing human capital formation, entrepreneurship and knowledge creation and diffusion. To many old industrialised regions what this meant was a
challenge to acquire and put in motion the prerequisite of modern location theories. This included overhauling of previous convergence measures and promote mechanisms the educational infrastructure as well as the way of thinking in terms of business development. To implement these theories into practice poses both threats and opportunities that have far-reaching implications on the role of human capital formation in general, and on higher education and research development in particular. In addition, regional development programmes focused on mechanisms to promote the growth of small- and medium-sized firms (SMEs). Encouraged by the debate on the new economy and the arrival of the information communications technology (ICT), regions focused on measures to induce knowledge creation and relationship building mechanisms in a hope to regenerate their local economies. Some areas are concentrating on accessing codified knowledge, while others are benchmarking successful concepts such as innovations system (Lundvall 1990), Clusters (Porter 1990) to improve the competitive strengths of their regions. The aim of these developmental initiatives is to overcome potential disadvantages of firms in terms of size and location and to facilitate learning, innovation, knowledge sharing, and creation of territory-specific types of knowledge that are crucial to competitiveness.

In this context, in the 1990s, the regional development debate in Sweden, for instance, was the three areas of northern Italy, Silicon Valley and, closer to home, the Gnosjö region, have inspired local and regional development policy initiatives. The clusters of the Third Italy (Becattini 1990) are spread over several districts of Italy (Johanisson 1995; Brulin 2000), Silicon Valley (Saxenian 1995) is one geographically large area, and the Gnosjö region of Sweden consists of four municipalities. Silicon Valley is argued to have all the necessary characteristics of the new economy. Its success was partly attributed to the birth of the micro-electronics industry and the abundance of venture capitalists, professional networks and institutions of higher education, which are characteristics that the other two regions lack. The industrial dynamics of the other two regions are instead characterised by the presence of small- and medium-sized enterprises and their ability to generate entrepreneurial spin-offs, to manage interdependencies and to develop horizontal relationships that involve both formal and informal institutions.

In addition, compared to Silicon Valley, the regions in Italy and Sweden are not known for being home to knowledge-based firms. Instead, these regions are known for the ability of the actors to engage in horizontal development dynamics in which relational assets play a vital role. In terms of regional development, the only similarity these three regions have in common is that they were situated “in the periphery” before they were discovered.
Also, the onset and proliferation of the digital technology was hailed as an important regional development tool that would compensate municipalities in peripheral regions for the lack of higher education institutions. By all means, ICT application has improved the competitiveness of firms and it has brought with it the emergence of new kinds of firms, such as call centre firms. These firms arrived at a time when it was observed that more and more economic activities could be outsourced or carried out in temporary projects by other companies. In a hearing at the end of 1990, on regional development, it was alleged that four out of five companies that had received state or regional subsidies to establish a presence in remote areas were firms in the ICT industry (SOU 1999:138).

In a nutshell, what we have witnessed the last two decades could be described as a concept driven regional development in which modern location theories with a point of departure in knowledge building and relational exchanges between different actors became the Mädchen für alles. According to modern location theory, product and process innovations take place through interaction between geographically proximate actors in systems like clusters (Porter 1990), innovation systems (Lundvall 1990) and competence blocs (Eliasson 1996). Through relational linkages facilitated by physical proximity, the diffusion and transfer of contextual knowledge has become an important mechanism to determine the competitiveness of not only individual firms, but also the location they function from. Consequently, collective knowledge spillover between different geographically proximate actors results in new products and/or services and the improvement of the existing ones (Audretsch, 2000; Maskell et al 1998; Porter 1990).

The concept of clusters (Porter, 1990) in particular, was on the lips of policy makers. As Porter (1998) argued the assumption was that the “global economy” does not diminish the role of the local environment but links production facilities to a local milieu with global strategic capabilities. Global firms need local capabilities to reach customers and they are dependent on locally produced resources to be able to compete globally. According to Porter (1998), this creates a paradox. Maskell et al (1998) argue that the carriers of knowledge are accessed locally and by developing their competences and operational capabilities the firms become able to innovate products and processes.

Also the concept of creative cities (Florida, 2002) should be seen from the above context where knowledge creation, innovations and creativity have come to be viewed as mechanisms that impact location of economic activities and the policy implications related to that.
References


Bergström, F. (1998) “Regional Policy and Convergence of Real per capital income among Swedish regions” SSE/EFI working paper series in Economics and finance no.8


Brusco, S. (1986) “Small Firms and Industrial Districts; The experience of Italy”, In New Firms and Regional Development in Europé; Edited by D. Keeble and E. Wever, London- Croom Helm”.


Mellander, C and Florida, R. (2007) The creative class or the human class”, Cesis, electronic working

Morgan D. Thomas. (1969) “Growth pole theory: An examination of the some of its basic concepts”. In “Hansen; Growth centres in regional development.


Sweden. Population and Environment, 3(2) (125-134).


Introduction

We live in a world where we get to know new innovations every day. The smartphone applications, computer software, daily habits, household appliances and modes of transportation are just a few to name. Perhaps, they have been there for a while, but it is until we adopt them that they become innovations for us. However, why does one innovation get adopted but not the other? This is a very important question worth pondering, especially for companies, governments and individuals that attempt to kick off the diffusion of particular innovations.

Diffusion often refers to how innovations propagate in a specific societal environment. Diffusion framework has been used to explain the spread of innovations, which is perceived as new by potential adopters (e.g., by individuals or organizations). The “newness” poses uncertainty and underscores the importance of diffusion study. Innovation and its diffusion often have positive effects on social, technological and economic development. However, innovation alone without the diffusion process would not have much effect (Hall, 2005).
Apart from this introduction, this book chapter is structured as follows. Section 2 presents the definition of diffusion. Section 3 explains the three components of diffusion: innovation, sources and adopters. Section 4 introduces two case studies (one from Germany and the other from Bangladesh) and explains how the diffusion theory can be applied in real life contexts. Last but not least, section 5 discusses managerial and policy implications, section 6 suggests some further readings and section 7 discusses the methodological considerations.

**What is diffusion?**

The definition of diffusion is “the process in which an innovation is communicated through certain channels over time among the members of a social system. It is a special type of communication, in that the messages are concerned with new ideas” (Rogers, 2003, p. 5). The seminal book Diffusion of Innovations, by Everett M. Rogers (1962), has been the core of a large body of literature in diffusion research. For decades, the theory has been used by many scholars from a variety of disciplines, such as economics, sociology, engineering, and management.

The diffusion is not just a passive process; rather, it involves a complex process with incremental adjustments needed to make all the parts of a system fit together (Nathan. Rosenberg, 1994). In fact, the process involves several related activities even before the actual diffusion takes place (N. Rosenberg, 1976). In some cases, the activities of individuals, companies or policy makers may result in some innovations that can be easily diffused by early and later groups of adopters. The socioeconomic cost of adoption at a later stage of the diffusion significantly becomes lower. Because, if the innovation continues to diffuse, technical and commercial risks will be reduced and production costs will decline.

Historically, diffusion has been an important element of ethnological research addressing how cultural innovations have spread among cultures. It also tried to shed light on whether specific innovations are the results of autonomous processes or dependent on diffusion. Academically, diffusion has been used in various branches of social sciences and has been given various definitions. Studies in diffusion research have been focused on explaining the variables that influence the adoption process. Diffusion and adoption processes are highly complicated, having several unforeseeable and context-specific factors. In addition, diffusion processes are almost always associated with something new, magnifying the degree of uncertainty and, hence, resulting in unpredictable outcomes. Therefore, studying diffusion, its pattern, its model, its factors, and its effects can help scholars and practitioners to understand
better the process and to formulate increasingly more efficient, fruitful, and productive innovation diffusion schemes.

When studying diffusion, it can be important to understand the innovation at hand. For example: Is it public or private? Is the cost high or low? Is it standardized or dynamic?

**Public vs. Private**

The diffusion of innovations might result in either private or public consequences, or sometimes both. If there are some public consequences at hand, the diffusion process usually involves several actors and collective actions, such as policy makers, organizations, and social movements (Wejnert, 2002). For example, environmental innovations have both public and private consequences, as they avoid or reduce environmental harms for the whole society. The diffusion of environmental innovations also involves policy makers, as it was seen in the case of the Europe 2020 strategy of European Union (EU). As a public consequence, it is widely assumed that the diffusion of environmental innovations will ensure future employment, contribute to economic growth in Europe, and respond to today’s major societal challenges, especially environmental (ETAP, 2010). In general, the diffusion of innovations that have public consequences is a more complex and lengthier process in comparison with those that have only private consequences.

**High cost vs. Low cost**

The cost of adoption of an innovation could be both monetary and nonmonetary forms, and it is directly or indirectly associated with the innovation (Wejnert, 2002). In general, for high-cost innovations, the economic aspects are often the most important factors for understanding the diffusion of such innovations (Rogers, 2003). When the cost of adoption of an innovation is perceived as high, the potential adopter needs to understand clearly the outcomes, benefits, risks, and costs, both in short-term and long-term, associated with the adoption. That is why high-cost innovations require more time to be diffused and involve more actors for the diffusion process in comparison with low-cost (non-cost) innovations.

**Standardized vs. Dynamic**

In some cases, innovations do change or get modified in the process of diffusion. The change can be so continuous that the innovation gets modified in every adoption that takes place in space and time. For instance, an innovation can evolve during the diffusion phase when the adoption is highly restricted by adopter’s settings. Some possible examples could be cornrows hairstyles, as adopted by individuals; solar PV systems, as adopted by households; and internet websites, as adopted by firms. However, some other
innovations can be standardized in one or a limited number of forms, appearing in the market on an incremental basis, e.g., automobiles, Facebook, or smart phones. In this case, the innovation does not represent a fully heterogeneous form in space and time, instead, it is an incremental one.

Three Components of Diffusion

In the history of science, diffusion as a term was primarily used in the chemistry and physics in order to define the process of movement of molecules from high-concentrated regions to other regions (for example Fick's laws of diffusion from 1855). But, can the diffusion of innovations conceptualized as similar to the diffusion phenomena in physics and chemistry? Perhaps to some extent, but not completely. In this book chapter, however, we refer to some analogies between diffusion of innovations (in social world) and diffusion of molecules (in physical world). For example, Figure 1 illustrates a three-component model of the diffusion of innovations: source (in analogy with high concentrated regions), innovation (in analogy with molecules) and adopter (in analogy with low concentrated regions).

![Figure 1. Simplified diffusion process (Sriwannawit & Laestadius, 2015)](image)

Innovation

Innovation is the core element of diffusion of innovation process. The assumption is that innovation can diffuse among potential adopters. Innovation is a very broad term and indeed can be anything as long as it is perceived new by adopters. The variety of innovations that are analysed in the literature is evident in the early scholars of diffusion. For example, Griliches (1957) analysed the diffusion of hybrid seed corn, which is a product of controlled crossing of specially selected parental strains among American farmers. Mansfield (1961) studied the diffusions of twelve types of innovations, such as a trackless mobile loader and a continuous mining machine among firms in particular industries.

The availability of a new technology or innovation does not necessarily motivate its adoption by individuals. The perceived attributes of an
innovation, which is contingent upon the adopters, explain 49-87% of the variance on the different diffusion rates of different innovations (Rogers, 2003; Tidd, 2009). These attributes are relative advantage, compatibility, complexity, trialability, and observability.

- **Relative advantage** refers to the degree to which an innovation is perceived to be better than the incumbent idea, technology, or practice and is usually expressed as economic profitability. However, non-economic factors (e.g., quality, satisfaction, environmental awareness and social prestige) are also important, as it is the case of the PV diffusion (e.g. Jäger, 2006a; Müggenburg, Tillmans, Schweizer-Ries, Raabe, & Adelmann, 2012; Palit, 2013). Greenhalgh et al. (2004) also emphasize that potential users will likely not consider the innovation if they do not see relative advantages, which are foremost measured in economic returns. Moreover, there are decisive social factors such as user satisfaction and prestige that influence an individual’s perception of the relative advantage of innovations.

- **Compatibility** is the degree to which an innovation is perceived as being consistent with the existing values (e.g., sociocultural values and beliefs), past experiences (e.g., previously introduced ideas), and the needs of potential adopters. Several empirical studies points to a direct relationship between the compatibility of an innovation and its adoption (McEachern & Hanson, 2008; Shum, 2013).

- **Complexity** is the degree to which an innovation is perceived as being relatively difficult to understand and use. Generally, there is an inverse relationship between the perceived complexity of an innovation and its adoption rate (Kai-ming Au & Enderwick, 1999; Labay & Kinnear, 1981; Völlink, Meertens, & Midden, 2002).

- **Trialability** is the degree with which an innovation may be experimented on a limited basis. Innovations with high trialability often have a higher diffusion rate (Makse & Volden, 2011; Rogers, 2003), although some other studies (Labay & Kinnear, 1981; Völlink et al., 2002) indicate an absence of a relationship between trialability and the adoption of innovations in the energy sector.

- Finally, **observability** is the degree to which the results of an innovation are visible to others. According to Tidd (2009), the rate of adoption of an innovation increases when it is easier to see the benefits of this innovation.

Even though these attributes have been widely studied and used in scholarly research, they have sometimes been considered as abstract and irrelevant for practitioners. These attributes are, in fact, simple to understand and can be operationalized in practical cases but it is crucial that the attributes be analysed as ‘perceived’ not as ‘intrinsic’ (Weiss & Dale, 1998).
Sources

Sources of diffusion could be the individuals, organizations or institutions at which the diffusion process originates. They can be innovators, government agencies, companies or any individuals. Sources are usually important change agents – that influences the decisions of potential adopters in a desirable direction- in the diffusion process. This means that potential adopters can be influenced to adopt an innovation by the pressure of the social system generated via adopters, public policies, shareholders and organizations (Bass, 1969; Frondel, Horbach, & Rennings, 2008). Nevertheless, change agents’ efforts and the adoption rate of a technology do not always correlate. More payoffs of change agents’ efforts occur at specific stages. The highest response to the efforts is when an opinion leader – a relatively well-known individual or organization that has the ability to influence the opinions of others – adopts the change.

In the literature, the common sources of diffusion are often conceptualized as policy measures, previous adopters and firms. In the following bullet points, we exemplify these.

- **Policy measures**: Policy makers can foster specific regulations to guide suppliers and adopters to choose an innovation or a specific design, which in return may increase the economical relative advantage of the innovation. For example, in the case of environmental innovations, the policy can have the capacity to create expectations for an innovation that is essential to foster the diffusion. In this context, until 2012 in Germany, policy makers provided clear signals regarding the growth potential of solar PV systems through the implementation of feed-in tariff (Dewald & Truffer, 2012; Hoppmann, Peters, Schneider, & Hoffmann, 2013; Staffan Jacobsson & Lauber, 2006). Such tariffs influence the perceived economic relative advantage of solar PV and, therefore, the diffusion (Jäger, 2006b).

- **Previous adopters**: In general, potential adopters are often influenced by what they see and hear from their peers. The peer effects has received a lot of attention from scholars analysing the academic and health practices (e.g. Trogdon, Nonnemaker, & Pais, 2008; Zimmerman, 2003). In the case diffusion of PV systems, several scholars (e.g. Bollinger & Gillingham, 2012; Graziano & Gillingham, 2014; Müller & Rode, 2013) have also asserted that the peers, who have already adopted the PV systems in the same neighbourhood, increase the diffusion rate among potential adopters. This resonates quite well with the recent work of Pentland (2014) on social physics. He emphasizes that innovations and good ideas spread faster if the potential adopters have regular physical, e.g. face-to-face, interactions. In a well-connected global world, although the potential PV adopters
might have tight virtual connections with their peers from other regions, the most of physical interactions takes place in the local neighbourhood. As Graziano and Gillingham (2014) also argue, such spatial dimension is especially critical for innovations that have both private and public good characteristics, e.g. solar PV systems.

- **Firms:** The primary economic function of an industrial firm is “making use of productive resources for the purpose of supplying goods and services to the economy in accordance with plans developed and put into effect within the firm” (Penrose, 2009, p. 12). In the renewable energy sector, the activities of firms not only result in supplying goods and services but also in the creation of new knowledge and the development of different types of designs (S. Jacobsson & Bergek, 2004). The firms are important stakeholders in the development of renewable energy policies, and therefore, they try to influence the political decisions about the design of financial support systems and the grid access (Wüstenhagen, Wolsink, & Bürer, 2007, p. 2686). For example, in Germany, the energy utility firms are well-known for their effective lobbying strategies for renewable energy, such as, regular and personal maintenance of contact to politicians and forming a policy network through associations (Sühlsen & Hisschemöller, 2014). Firms could be conceptualized as change agents, which influence the opinion of potential adopters. For the meso/macro level analysis, firms could be conceptualized as actors or complementary inputs. Whether the analysis is at micro or meso/macro level, local solar firms are often identified as important drivers of the diffusion of solar PV systems. For example, in the case of PV systems in Germany, Dewald and Truffer (2012) identified that local solar firms stimulated rapid diffusion and market formation. They argued that the successful diffusion is not only driven by the strong policy support and favourable geophysical conditions, but also by the market formation activities of local solar firms. This is in line with the study of Fabrizio and Hawn (2013) which analysed the role of local solar companies in the USA. Conceptualizing the local firms as complementary inputs, they identified that local firms do function as important drivers of diffusion of PV systems.

### Adopters

Adopters are the individuals, organizations or institutions who take up or embrace the innovations. This means that adopters are the individuals or other units of adoption that make the decision to adopt the innovation (Rogers, 2003). Any individual can be a potential adopter, however, not all potential adopters necessarily decide to embrace the innovation at any certain point of time. 2 shows the points in time at which different groups adopt innovations. The distinction between the different adopter groups is based on behavioural characteristics regarding innovativeness and corresponding to
normal distribution statistic percentages. Innovators are those who want a certain product as soon as it becomes available, are willing to take risks, and have financial capabilities. They are cosmopolites who act and have contacts regionally and globally. Early adopters are a larger group who also seek new products but are less sensitive to “hype” (e.g., they look more into functionality). They are often local and generally have the highest degrees of opinion leadership (Rogers, 2003). The early majority is the first mass of people to adopt a product, and this is where the curve reaches maturity. The late majority adopts when the majority of the market is already familiar with the product. Sales tend to slow during this phase. Finally, the laggards adopt when the product is soon to be removed from the market. These adopters are more price-sensitive and sceptical (Rogers, 2003).

![Adoption curve and the chasm. Adapted from Rogers (2003) and Moore (2002)](image)

Typically, the adoption process begins relatively slowly, but once a critical mass is reached, it becomes an automatic mechanism that forms an S-shaped curve. This critical mass is one reason that, after a relatively slow start, the rate of adoption can form an S-shaped curve. During the diffusion process, influence can be exerted between adopter groups. Individuals who are influential within the social system and who spread information about an innovation are defined as opinion leaders and are generally to be found among the early adopters (Rogers, 2003).

According to Moore (2002), there can be gaps between different adopter groups. One of the most important gaps is between the early adopters and the early majority, defined as the chasm (see Figure ). This occurs when a new product or service cannot be translated into a significant benefit (Moore, 2002). Sometimes, early adopters can create bad references for the early majority. For example, StudiVZ of Germany (founded in 2005, a pioneer social networking platform – similar to Facebook) has failed to cross the chasm between its early adopters (in Germany, Switzerland and Austria) to the early majority (worldwide individuals elsewhere). However, on contrary,
Facebook (founded in 2004) has successfully crossed the chasm between its early adopters (the students in the US colleges) to the early majority (worldwide individuals elsewhere).

The diffusion rate of innovations, which is the ratio of current number of adoptions to the total number of potential adoptions, is mainly based on the attributes of an innovation as perceived by potential adopters (as discussed previously). Besides the attributes of innovation, there are some other factors that can affect the adopters. Such as:

- **Types of innovation-decision**: affect whether an innovation should be adopted or rejected. They can be optional decision made by individuals independently from other members; collective decision reached by consensus with others; and authority decision made by relatively few members who possess status, power or expertise. These three types are arranged on a continuum indicating the power that each individual possesses in decision-making – from complete power via partial power to no power at all. The rate of adoption decreases when there are more people involved in the decision process. However, the diffusion process may involve a mix of all of these decision-making types, depending on the type of technology, regulations and adopters, as is sometimes the case of the renewable energy technologies in different countries (Bodas-Freitas et al., 2010; Reardon, 2009). For example, if there is an innovation to build on an apartment, multiple adopters (who live in the same apartment) may have a say on the final decision.

- **Communication channels**: affect the innovation-decision process. Communication is the process through which information is created, received, and shared. A main goal is to achieve mutual understanding among the participants. Interpersonal communications (including non-verbal observations) and mass media channels (television and internet) are important influences on the diffusion rate of the innovations in a social system. The former is more powerful in convincing a social system to accept a new innovation (Mahajan, Muller, & Bass, 1990; Rogers, 2003). Communication between adopters and the observability of the adoptions can induce peer-effects, whereby the decision of potential adopters may be influenced by the previous adopters (Bollinger & Gillingham, 2012). Recent literature has paid much attention to how peer-effects influence the diffusion of PV technology (Müller & Rode, 2012; Rode & Weber, 2012).

- **Nature of the social system**: can affect the adoption. A social system is a group of interrelated actors. Rogers (Rogers, 2003, p. 24) defines a social system as “a set of interrelated units that are engaged in joint problem-solving to accomplish a common goal”. The members of a social system may be individuals, informal groups, organizations and/or subsystems. Units in a system can differ in their behaviour by
means of homo- and heterophily: homophily refers to the similarity between individuals, e.g., regarding education level, beliefs and social status, whereas heterophily is when individuals differ on these attributes. One distinctive challenge is that individual adopters are often heterophilous. This makes it difficult for an actor attempting to diffuse an innovation to choose only one single approach (Rogers, 2003).

Potential adopters can be influenced to adopt an innovation by the pressure of the social system generated via adopters, public policies, shareholders and organizations (Bass, 1969; Frondel et al., 2008). Some recent research have identified the effects of network externalities as being significantly important for the diffusion rate of innovations (Goldenberg, Libai, & Muller, 2010; Van den Bergh, 2013).

The adopters are crucial in shaping and diffusing technologies. For example, collaboration of users in energy technology can lead to greater diffusion, technological development, and possibly new product innovation (Ornetzeder and Rohracher 2006). Nevertheless, since the 1980s, when there was an attempt to implement renewable energy technology, social acceptance has been neglected. Initial surveys of public acceptance of renewable energy, in fact, showed very high support of renewable energy. However, in reality, a large difference exists between general public support and support for specific projects or implementations. Policy makers have been misled by this discrepancy between general positive support and low support for specific contexts directly affecting individuals. Moving from general to specific and from the global or national to the local level, direct investigation of public support of local sites should be investigated. Acceptance of general projects and resistance to specific implementation may occur because people tend to support renewable energy as long as it is not in their backyards. Therefore, specific cases must be investigated to provide understanding of local contexts (Wustenhagen et al. 2007). The inclusion of users in technology diffusion research can provide better understanding of environmentally friendly technologies and enhance dissemination. Recently, there has been an interest in academia in investigating social acceptance and users’ perspectives through the case-study approach (see e.g. Rohracher 2003 in Austria; Mallett 2007 in Mexico; Muggenburg et al. 2012 in Ethiopia; Shyu 2013 in China). For small-scale renewable energy technologies, like household PV systems, Wustenhagen et al. (2007) suggested diffusion of innovation literature (Rogers 1995) can help in understanding social acceptance because it can be regarded as market acceptance or a market adoption process. Nevertheless, because of the nature of energy technologies, which are constrained to other infrastructures, their diffusion can be more complicated than that of other products.
The roles of potential users are of high importance for all four elements of the diffusion: innovation, communication channel, time, and social system (Rogers, 2003). First, in terms of the innovation itself, understanding users’ needs is a vital determinant of its success. Users’ involvement in the innovation process leads to development of a more valuable innovation, enhancing its adoption. Therefore, integrating users from an early stage and actively engaging them throughout the process are necessary. Second, an appropriate communication channel must be selected for effective facilitation of the diffusion process. Along with the communication channel, the cognitive distance between the source and the recipient and users’ absorptive capacity should be taken into account. Third, the time and its three sub-aspects are strongly connected and highly dependent on users. The first sub-aspect, the decision-making process, involves users’ deciding whether to adopt or reject the innovation. The second aspect is the relative earliness or lateness of its adoption by other members in a social system. The last aspect is the rate of adoption, measured directly by counting the number of users who choose to adopt the innovation within a certain time frame. Finally, users are obviously the main part of the social system. Without users, the system would not be complete because the function of the system would not be fulfilled.

**Empirical cases: Solar Photovoltaic Systems**

In today’s world, we are facing several environmental and societal challenges at which understanding diffusion of innovations could be a part of the solutions. For example, concerns about climate change and limited resources of fossil fuels have prompted governments to support the emergence and diffusion of renewable energy systems. At the same time, there are still 1.3 billion people living without electricity (IEA, 2012). Generating electricity from solar energy, an abundant and renewable source, using photovoltaic (PV) systems is one means to tackle these challenges.

A new technology’s being superior to the preceding technology does not guarantee its successful dissemination, as is the case for PV systems (Miller 2009). Photovoltaics (PV) is an electricity generation method using solar radiation. Based on grid connectivity, PV systems can be separated into two categories: on-grid and off-grid. An on-grid PV system is integrated into the utility grid. During the day, electricity is generated via the PV system. The excess amount is transferred to the national grid for general use at any grid-connected location. When the PV system cannot generate sufficient electricity, the electricity can be taken from the utility grid; thus, a battery is not a necessary component. On the other hand, an off-grid PV system is used in remote areas where the grid is unavailable; hence, a battery is an important part of the off-grid system. During the day, solar energy is collected and transformed into electricity. The excess amount of electricity is stored in
batteries for later use. Despite the use of PV technology for rural electrification since the 1960s (Lorenzo 1997) together with its rapid technological development and price decline during the last few years, its use is still low compared to conventional energy sources (REN21 2014). The rapid price drop has made PV systems an attractive energy technology, even among low-income inhabitants. However, the significant price decline has not been fully appreciated by relevant actors, such as policy makers and users who still perceive a PV system as an expensive option (Bazilian et al. 2013).

In the following two sub-sections, we present two case studies. One of them is about off-grid solar PV Systems in Bangladesh, while the other is about on-grid solar PV systems in Germany (see Table 1).

<table>
<thead>
<tr>
<th>Case Study</th>
<th>Innovation</th>
<th>Main initial source</th>
<th>Adopters</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>Off-grid solar PV Systems</td>
<td>Grameen Shakti</td>
<td>Some local Bangladeshis</td>
</tr>
<tr>
<td>B</td>
<td>On-grid solar PV systems</td>
<td>Hartmann Energitechnik GmbH</td>
<td>Some local Germans</td>
</tr>
</tbody>
</table>

These two cases from two different continents have been selected to represent the diffusion process of PV systems because of three main reasons. Firstly, the underlying technology are similar (solar PV systems) but the applications are different (on-grid vs. off-grid). Secondly, the sources in both cases are of the same type (i.e., private firms) but they undertake different firm strategies. Thirdly, the adopters in both cases are householders, however with different socioeconomic characteristics. Overall, all these differences and similarities between two cases let us explore how vital the characteristics of innovation, source and adopters are to the diffusion of innovations.

**Case study from Bangladesh**

This case study is mainly based on Grameen Shakti, a local energy company operating in rural Bangladesh. Bangladesh is an overcrowded country with approximately 160 million inhabitants. One-third of the population lives below the poverty line (CIA, 2013). More than half of the rural population do not have access to electricity (IEA, 2014). In addition, power services in rural areas are of poor quality. There is a great need and a large market for reliable
electricity at an affordable price. This is a potential for a stand-alone off-grid energy technology like Solar Home System (SHS), which is a set of photovoltaic systems installed on the rooftop to convert sunlight into electricity. It is a self-contained generation and distribution system that generates much lower maintenance and operating costs than fossil fuel alternatives (Cabraal, Cosgrove-Davies, & Schaeffer, 1998).

Case study in Bangladesh analyses the diffusion of energy technology for rural electrification. It is a highly relevant and timely challenge for many communities in developing countries. We focus on the diffusion of SHS in particular among the population with low income or those belonging to the “bottom of the pyramid” (BOP). Strong focus in the BOP discourse has been placed on international business or multinational companies (Prahalad, Di Benedetto, & Nakata, 2012; Prahalad & Hammond, 2002) but it tends to neglect the role of government (Bawakyillenuo, 2009) and local entrepreneurs or industrial activities within developing countries (Linna, 2012).

Grameen Shakti was founded by Muhammad Yunus, also the founder of Grameen Bank and recipient of the Nobel Peace Prize for his microfinance concept. In Yunus’s opinion, development must be based on a positive economic change in the bottom half of the population (Muhammad, 2003). Consequently, Grameen Bank as well as Grameen Shakti focus their activities on the BOP (Prahalad, 2010). GS was established to disseminate affordable and renewable energy technologies to rural and remote areas in Bangladesh. The company has become one of the largest of its kind in the world (Asif & Barua, 2011; Sovacool & Drupady, 2011) with approximately 1,500 offices and 12,000 employees covering all 64 districts of Bangladesh. The diffusion of GS’s first SHS began in 1996. By the end of 2013, GS had provided SHS to more than 1.3 million rural households across the country (GS, 2010). SHS has successfully penetrated the rural Bangladeshi market with no direct subsidies (Kolk & van den Buuse, 2012). On the contrary, SHS is still an emerging technology with low dissemination in most countries (Rebane & Barham, 2011). This motivates us to learn more regarding the factors that affect the high adoption rate in Bangladesh, which we discuss in the following.

Miller (2009) identified two types of capital that are barriers to entrepreneurs dealing with solar energy. First, market infrastructure capital—such as local branches, sales people, and technicians—is necessary. Second, customer finance is important in order to make the products affordable for the target group. Grameen Shakti has overcome these two barriers, and its diffusion is at a large nationwide scale. The analysis here focuses on the characteristics of the source that are specific to the BOP market, here identified as a social entrepreneur. These source characteristics that equip Grameen Shakti with competitive advantage comprise four items: (a) unique company background in development work related to the source’s history, leader, and values; (b) good qualifications and strong effort of change agents highly involved with
local communities, both before and after the adoption; (c) affordable and manageable economic schemes specifically designed for the low-income adopters; and (d) a distribution system quick to respond to adopters’ financial situations. The source characteristics are assumed to be the most important factor affecting the high adoption rate of PV systems in Bangladesh because they are unique and they specifically address the BOP market. Understanding the source characteristics can equip various types of actors with the capabilities to begin a diffusion process in the low-income segment. The Grameen Shakti case reveals that a local social entrepreneur can drive the diffusion process by building a previously unattractive market to become a market with large potential adopters. The social characteristics of Grameen Shakti may be largely because of its founder, Muhammad Yunus, and his ‘innovation’ of the microfinance concept. As opposed to Prahalad’s BOP concept, which may be considered a top-down approach, Grameen business may be considered a complementary perspective, focusing on capability building in local firms rather than waiting for investment from multi-national companies.

In summary, this case study shows that the source could be conceptualized as a social entrepreneur. The source characteristics may be the most important factor that affect GS’s adoption rate because they are unique and specifically address the BOP market. This market has great potential but also contains vulnerable consumers. Social entrepreneurship is one manner in which to approach this market. GS’s strategies have overcome a common barrier of high initial cost. The GS case reveals that a local social entrepreneur can drive the diffusion process by building up from a previously unattractive market to become a market with large potential adopters.

**Case study from Germany**

Germany is a leading producer of solar photovoltaic-power, accounting for one-third of the global photovoltaic (PV) systems’ installed capacity (GTAI, 2013). Although Germany has a relatively limited solar radiation potential, it has outperformed countries that have a larger land area and higher solar radiation, such as France, Spain, Turkey, Brazil and Japan. The mechanisms behind the German story of PV systems’ diffusion have been discussed at length in the literature (Dewald & Truffer, 2011; e.g. Staffan Jacobsson & Lauber, 2006; Karakaya, Nuur, Breitschopf, & Hidalgo, 2014). One of the most important drivers of the diffusion process has been the activities of local solar companies (Dewald & Truffer, 2012). According to some studies (Lettner & Auer, 2012; Pérez, Cervantes, Báez, & Domínguez, 2013; Spertino, Leo, & Cocina, 2014), solar PV energy in Germany has already achieved grid parity by 2012, i.e., solar PV energy can directly compete with conventional electricity sources in terms of the levelized cost of electricity generation.
The German feed-in tariff scheme is widely accepted as the strongest driver for the diffusion of PV since 2000 (Dewald & Truffer, 2011). This scheme ensures that solar PV adopters (when they supply electricity to a grid) are paid with fixed feed-in tariffs over 20 years, beginning from the time the solar PV systems are connected. However, the feed-in tariff for solar PV systems has decreased more rapidly than that for any of the other renewable energy technologies (Wirth, 2013). Although solar PV systems in Germany are often assumed to be at grid parity, the PV market has recently faced uncertainties related to the cuts in the feed-in tariff.

The focus of the case study is the source of innovation, Hartmann Energitechnik GmBH (HET) in Germany. This company has been chosen for two reasons. Firstly, it is located in the spatial lead market of PV systems’ diffusion in Germany (Karakaya et al., 2014). Secondly, it has witnessed the diffusion of PV systems since 1990s as being one of the pioneer local solar companies in Germany (Karakaya, Hidalgo, & Nur, 2015). It has been a part of solar initiative movement (Drücke, 2004) and featured in several energy magazines (SB, 2007; e.g. SWW, 2004).

The company offers not only PV systems, but also some other technologies, including solar thermal systems, pellet ovens etc. As shown in it is located in a rural area, near a village, called Oberndorf in Rottenburg (Am Neckar). It is a local solar company founded in 1995 by a local entrepreneur, Thomas Hartmann, a native of Oberndorf. He is also the co-founder of two solar initiatives: Solar-Partner e.V and Sonnenhaus-Institut e.V. The former is a network of companies, freelance solar consultants and partner companies. The latter is an association of architects, engineers and managers of the solar industry, focusing on solar-heated and solar-electrified buildings.

This case study demonstrates several important motivators for adoption. Achieving grid parity does not necessarily motivate potential adopters for PV systems. As discussed by Rogers (2003), for wide adoption, innovations should not only have high compatibility, high trialability, high observability, high relative advantage and low complexity but also be communicated and driven by change agents, e.g. adopters and other actors.

Most of the adopters in our case have been motivated for PV systems in order to be self-sufficient and independent from electricity supply. Such desire is often complemented with environmental awareness, peer effects and financial stability. This is in line with previous research, which investigated similar phenomenon in other countries before grid parity (Balcombe, Rigby, & Azapagic, 2013, 2014; Jäger, 2006a). In our case, the increasing electricity retail prices, as influenced by policy measures, have motivated potential adopters to be less dependent on electricity supply. In this context, PV systems have been often perceived to be compatible with the desire of potential adopters towards being self-sufficient and independent.
Moreover, complementing the previous research (Dewald & Truffer, 2012; Fabrizio & Hawn, 2013), the local solar company is identified to be an important motivator for adoptions. The vision of the company and its local entrepreneur has reduced the perceived complexity of PV systems. Several activities organized by the local entrepreneur, eg., solar-walks and open-door days, have also increased perceived trialability and the observability of PV systems. Because PV systems require some level of knowledge about the technology and its operation, a high level of communication between local solar companies and adopters is a key factor to minimize the perceived complexity and to facilitate the decision made by the adopters.

**Managerial and policy implications**

The diffusion of innovations in general and solar photovoltaics in particular have several managerial and policy implications. This book chapter provides two examples focusing on the source of innovation. The implications are drawn from previous studies and two cases: one from a developing country and the other from a developed country. The case studies discussed in this chapter suggests that the involvement of all stakeholders—adopters, local communities, firms, international organizations, financial institutions, and government—are crucial when it comes to high-cost dynamic innovations with public consequences. Without proper collaboration, effective marketing, and dedicated government support, the barriers to adoption cannot be easily overcome. However, technology diffusion is context specific and so is the adoption of PV systems. Therefore, in general, the local conditions of a particular case of an innovation should be well understood by companies, governments and individuals who are in favour of wide diffusion.

**Suggested further readings**

Diffusion research originated in and has been conducted largely by scholars in the advanced parts of the world. Various topics related to the diffusion of innovation have been studied since the 19th century (Boas 1896; Tarde 1903) and continue to be studied in academia. The framework is highly interdisciplinary in character. To date, this field has been addressed in a large amount of literature spanning such fields as folklore, cultural change, economics, technological change, and so on (Sriwannawit and Sandström 2015). Such depth and breadth of interest in diffusion research may be because of its wide application, the existence of many issues needing resolution, the specificities across cases, and the degree of generalization, among others.
A number of other studies can be considered complementary, filling in gaps in the diffusion framework. Therefore, we suggest the following studies as further readings:

- Cohen and Levinthal’s work on absorptive capacity (1990) addressed the importance of users’ capability in the exploitation of new knowledge based on related prior background. This work implies that adoption requires appropriate knowledge and some form of training. Their concept has been widely applied at the firm and organizational level.
- Nonaka and Takeuchi (1995), together with Jaffe, Trajtenberg, and Henderson (1993), help to fill in this complex understanding of the creation, utilization, and diffusion of knowledge, which is an intangible matter. This intangible aspect of diffusion may be lacking in Rogers’ work.
- The work of Wejnert (2002) who categorizes the variables that influence diffusion in three major groups: environmental settings (such as geographical and sociopolitical context), the characteristics of innovations, and the characteristic of adopters.

**On methodological considerations and the use of sources in this chapter**

It should be noted that this book chapter (i.e., 6 sub-chapters) are mainly based on the following research papers and dissertations:

Therefore, the content of this chapter may contain identical parts/paragraphs with the sources mentioned above.

**References**


http://doi.org/10.1080/13662716.2011.561028


http://doi.org/10.1016/j.ecolecon.2007.08.016


http://doi.org/10.1093/jeg/lbu036


7: Technological Transformation in Process Industries
Michael Novotny

This chapter is about a large-scale process industry where innovations have characteristics that are not well represented in the innovation management literature and where the transformation is related to technology shifts as well as changing business models. This industry located in the Northern hemisphere is trying to transform due to new market conditions, and to introduce new bioproducts based on woody biomass. In the following analysis inspired by the Swedish economist Erik Dahmén, I introduce the concept “diverging innovation” to the established toolbox of innovation analyses, arguing that this concept in general is useful for the analyses of process industries, and – in particular for the development block of wood-based biorefining – i.e. a set of diverging innovations of non-assembled products resulting in a renewal of industries. I will argue that understanding the nature of innovation in process industries have profound policy and management implications.

Introduction

When industries transform in a radical or disruptive manner, there are complex processes going on at different levels – i.e. old capabilities, technologies and markets get obsolete, new ones emerge and integrate with or displace the old. The unit of analysis for this study is the nature of innovation and transformation of pulp and paper related industries (PPI). This is a large-scale industry where innovations (i.e. product/process innovations) have characteristics that are not well represented in the innovation management literature and where the transformation is related to technology shifts as well as changing business models and industrial borders. This transformation is

50 Acknowledgement: This research has been funded by grant 12/09 from the Lars Erik Lundberg Foundation for Research and Education and by grant 2014-06082 from Vinnova, the Swedish Innovation Agency. I also thank my research colleagues of this book project on Industrial Dynamics at the SID division at INDEK, KTH (Royal Institute of Technology) for their comments on seminars, and particularly prof. emeritus Staffan Laestadius for his valuable comments, on earlier drafts of this chapter.
strongly linked to what is here labelled *diverging innovations of non-assembled products*. A diverging innovation is partly a diversification of an industry, but there is more to it. It is also a phenomenon of new sets of related products/processes that may replace industries of non-assembled products under certain conditions. First of all, to displace large-scale non-assembled fossil/petro-chemical or polluting products such as cotton with uniform product characteristics, there is a need for upscaling of biomass based substitutes holding the same chemical, molecular (e.g. carbohydrates such as bioethanol) or similar material composition – e.g. cellulose based plastics and (nano) cellulose fibers displacing fossil based polymers, plastic packaging, steel constructions, and the like. Since the production systems of process industries are closely connected to product specifications, innovation activities also are strongly affected. Processes and products are so tightly coupled that expressions like “the process is the product” have been coined (Linton & Walsh, 2008; Rosselle, 2012; Storm, Lager and Samuelsson, 2013). Biorefining or pulping, which is the main case of diverging innovation in this study, does therefore require upscaling of non-assembled biomass based products in order to partly substitute the huge amount of “fossils” in oil, chemical, construction, steel and transport sectors. A more profound discussion and analyses of the concept *diverging innovation* follows in section 4.

The context for this study is that of old industrialized countries in the Northern hemisphere – i.e. forest industries located in Northern European and North American countries endowed with “taiga” forests – mainly pine and spruce forests – as dominant raw material and to a large extent industries still active on declining markets (e.g. printing and graphic paper segments). These industries are trying to innovate and transform due to new market conditions, and to introduce new bioproducts (biomaterials, chemicals and fuels) based on their traditional natural resource and capabilities. A few fundamentals about pulping technologies and their main characteristics and differences are clarified in box 1.
Box 1. Pulping technologies

The large difference is between chemical and mechanical pulping. Chemical pulping is the process of converting wood (wood chips and saw mill residues) to pulp and separating the main components cellulose, lignin, and to some extent hemicelluloses from each other. In addition to the two most common chemical processes—kraft and sulfite—there is also a family of mechanical pulping processes. Mechanical pulping, of which thermomechanical (TMP) pulping is the most dominant, is not suitable for multiproduct biorefining. Neither the lignin nor the hemicelluloses are separated from the cellulose fibers, which means that mechanical pulps are both weaker and less ‘pure’ than chemically treated fibers. TMP is more suitable for printing papers. In addition, the energy needed to produce mechanical pulp and paper is much higher than the chemical process. (Biermann, 1996; Holik 2006; Novotny & Laestadius, 2014)

About 75% of the pulp mills in the world are constructed for chemical processes (70% kraft pulping and 5% sulfite), which makes them potentially suitable for biorefining. Chemical pulping technologies can develop organic products from the by-product called black liquor (approximately 50% of residues after chemically processing cellulose fibers). The sulfite process uses different solvents in the digester (cooking plant) than the kraft process to separate the wood compounds – acidic solvents instead of alkaline solvents – which lead to different functionalities of cellulose and lignin compounds. Sulfite is hitherto the technology that has developed most biorefinery products from wood – 7 millions of tons of textile pulps, almost 1,5 million tons lignosulfonate, a couple of hundred thousand cubic meters of bioethanol on the global market – and is the only process that today is able to separate a part of the hemicelluloses and develop ethanol derivatives (high value biochemicals). However, Swedish R&D organisations and Finnish companies have, based on the kraft process, recently commercially introduced technologies for extracting pure lignin (e.g. figure 3 and table 1) (Backlund & Nordström, 2014).

The analytical and conceptual toolbox applied to describe and analyse this industrial transformation is inspired by the Dahménian concepts “development block” and “structural tensions” (Dahmén, 1950 & 1970). The Swedish economist Erik Dahmén was strongly influenced by the Schumpeterian approach to innovations (Schumpeter 1934/1981). To Schumpeter’s entrepreneurial approach, Dahmén added a structural approach analysing industrial and technical change (cf. Laestadius, 2016), “through time within and among micro entities” (Dahmén 1989, pp 110). His approach still contributes with sharp and useful tools for policymakers as well as for industrialists and entrepreneurs/innovators for the understanding of industrial transformation.

In the analysis that follows – on the industrial and technological transformation from pulp and paper into biorefining industries – I will introduce the concept diverging innovation of non-assembled products to the established toolbox for innovation analyses, arguing that this concept is useful for the analyses of process industries in general, and this peculiar development block – wood based biorefining and pulping technologies – in particular. I will argue that understanding the nature of innovation in process industries have profound policy and management implications.
The chapter is structured as follows. In the next section I introduce theoretical concepts for understanding transformation over time, within and across industry borders. In section three there are examples drawn from the pulp and paper related technologies – of the rise and decline of papermaking and the trajectories for a biomass development block. Examples of the diverging innovation, of delineation of a biomass development block, its production chains of pulping and biorefining, and its mainstream technologies are presented here. This is followed by section four, a more profound discussion, on the nature of diverging innovations of non-assembled products and its consequences. Finally, I conclude with academic, policy and management implications and considerations for further research followed by recommended readings and a short description on methodological issues.

**Theoretical concepts**

This chapter is primarily conceptual. Using and developing the toolbox from Dahmén’s structural analyses on transformations, as well as established analytical tools from innovation management theory, the double aim is to increase the understanding of transformation, and innovation processes in process industries in general, the pulp and paper related industries in particular and to develop the toolbox for the analysis of innovations in process industry systems. After a short general methodological reflection, a theoretical overview and discussion follows in section 2.1 below.

**Industrial transformation and Dahménian dynamics**

According to well established innovation theory (Dosi, 1982; Utterback 1994; Laestadius 2000) process industries are often rooted in long ago developed technological regimes and have historically been characterized by gradual knowledge accumulation, irreversible investments, cognitive and physical inertia as well as high barriers to entry (and exit). As illustrated by Utterback (1994) the transformation to an integrated process has often been slow and stepwise and connected to huge capital spending. Heavy industries such as PPI are therefore characterized by a strong momentum towards exploitation of existing capabilities, very much due to a rather inflexible plant layout (Novotny & Laestadius 2014; Abdulmalek 2006; Nelson, 1994; Utterback 1994) and organisational (Henderson & Clark 1990, Tushman & Andersen 1997) as well as cognitive inertia (Laestadius, 2000).

As regards the industrial dynamic perspective on Schumpeterian innovation Dahmén (1950/1970) noted that new products, new methods, new forms of organization, and new markets – i.e. those forms of innovation Schumpeter (1934) lists in his *Theory of Economic Development* – can destroy the old and still not be detected in the analytical aggregates used by business cycle
theorists. The Dahménian approach to transformation instead identifies the interplays and interdependencies between industrial efforts and technologies, i.e. the causal mechanisms for industrial and technological transformation. This transformation usually takes place in the tensions and imbalances that occur between the agents of a development block (Laestadius, 2016). The imbalances occurring create necessities to adapt as well opportunities to grasp for agents involved. Agents may here be defined as the set of firms and their technologies which are complementary in the development block (Dahmén, 1950).

Depending on the character of technology shifts development blocks may change character over time. When analysing a development block the purpose is to identify innovations, significant R&D or industrial investments around which complementary – and sometimes competing – industrial and technological activities are formed over time. The reason is that these innovations – e.g. diverging biorefinery applications – may create new complementarities and new dependencies between specific functions or characteristics within the production system, across industry borders or between production and infrastructure. From these complementary mechanisms new technologies, new business models and new industries may be identified as well as old technologies, industries, etc. that are declining or closed down (Enflo, Kander & Schön 2008). Dahmén expressed that events like these may result in a two-way causality between innovations and industrial transformation. This methodological approach is brought forward in this chapter with the intention to grasp and identify industrial transformation and connect to the conceptual toolbox of Dahmén.

Dahmén’s focus – which has a family resemblance to Hughes’s (1992) concepts of salient and reverse salient within a technological system – is on the dynamics in vertical and structural linkages (i.e. inter-industry dynamics and large market/technology shifts), whereas Schumpeter’s description of transformation is more characterised by horizontal linkages (intra-industry dynamics) (Dahmén 1970; Hirschman 1981; Basalla 1988; Andersen 2011, Laestadius, 2016).

The concept of complementarities, which refers to how industries and technologies support and involve each other in order to accomplish or hinder transformation is closely related to structural tensions and how opportunities and necessities drive transformation (Dahmén, 1988). Transformation is, however, not a smooth process but rather characterized by structural tensions and time consuming industrial or institutional “bottlenecks” across industry borders that may enable – in case of successful investments, market/new raw material introductions, access to R&D infrastructure and skilled labour, etc. – or hinder further efforts – in case of complex regulations, malinvestments, delimited home markets, etc. Dahmén exemplified this by the electrification of Swedish industry in the early 20th and where the Swedish company ASEA
held a central role. It invented large power stations, but run into imbalances because of a weak home market. Only by financing and supplying industries as well as starting up subsidiaries based on electricity, ASEA accomplished the electrification of Swedish industry and strongly contributed to hydro power development in Sweden (Dahmén, 1950, pp 66-69).

Even more striking, according to Dahmén, were the structural imbalances in the technological and organisational field in fulfilling the development block. These are generally hard to distinguish from institutional and industrial bottlenecks but often resulting in more serious economic drawbacks. Dahmén found examples from companies involved in several production and/or distribution steps in the “chain”. Often a step in the chain with insufficient inputs, raw materials, or rigidities in transport techniques or market distribution, stopped the invention to become an innovation. Technological progress in a middle step could therefore be inhibited if not any progress were accomplished in the other steps (Dahmén, 1950, pp 69-71). A similar bottleneck occurred in cases when the different steps of the value chain were scattered among several companies. In these cases, an innovation could face bottlenecks in market introduction as well as in the distribution chain.

**Industrial transformation, shifts of pulping and biorefining**

In this section a presentation of the new industry conditions for pulp and paper related industries is introduced. This is followed by a subsection with the dynamics of technology waves, shifts of pulping, papermaking and biorefining, i.e. a thick description of an established technology platform and infrastructure for diverging innovations of non-assembled (bio)products.

As any mature industry the forest industries have gone through many transformations and structural changes. Historically, forest industries in the Northern hemisphere (North America and Northern European countries) have been rather successful by diversifying and gradually moving forward in the value chain in order to avoid market decline or raw material shortage. They went from timber exports and charcoal in the 17th-18th century to sawmill products in the 19th century. Later they developed into pulp and papermaking by using sawmill residues instead of cotton rags (i.e. primarily wood chips and top branches of round wood), and in the early 20th century for production of graphic papers. The new chemical pulping and bleaching technologies were invented in Scandinavia in the 20th century which later grew in scale and scope, particularly the new hygiene products and paperboard products expanded fast in the mid-20th century (Hylander 2009; Biermann, 1996; Holik, 2006, Toivanen 2012)
Since the end of the former millennium the pulp and paper industries located in the Northern hemisphere are confronted by manifold challenges. These influences or as Erik Dahmén (1970) would have labelled them; structural tensions or transformation pressures, have grown in the 2000s, i.e.: 1) Digitalization and multimedia; 2) Globalization and emerging economies; 3) A sustainability dimension with the aim of substituting fossils. While point 1 is a case of new technologies and markets replacing the old ones, point 2 and 3 are cases of pressures for a scramble for biomass resources, which is perceived in several parts of old industrialised nations, e.g. the Canadian West Coast exporting processed wood products/bioenergy products to Asia, as well as many parts of the EU countries with high ambitions with increasing the share of renewables, e.g. wood pellets, in the energy mix or as input in former petro-chemical products (fuels, chemicals, plastics, etc.).

From end of WWII to 2009 global paper and paperboard growth followed the worldwide economic growth closely – on average 3.4% growth/year (Hylander 2009). As of 2009-2010 this correlation was broken with annual growth rates being close to zero between 2010 and 2013 (Rennel, 2014). Many paper segments have declined significantly (except for tissue and paperboard) – particularly the decline in graphic and writing papers has dramatic figures. Between 2000 and 2009 newsprint demand was more than halved in the US (Hylander, 2009). During 2007-2012 over 80 paper mills were closed in North America (Pöyry, 2012); in the EU a similar number of closed paper machines (not mills like in North America, but with each paper machine producing on average between 100 000 and 250 000 tons annually) could be added in the last decade. However, chemical pulp production has increased gradually due to market demand in Asia, favourable conditions for paperboard demand and the growing niche, of diverging innovations of biorefining fields (textile pulp in primis) that will be presented more closely in section 4.1. In particular Brazilian, Finnish and Swedish kraft (chemical pulp) mills have increased production of pulp and diverging, additional “by-products” in recent years and also in coming years. In early 2016 Nordic pulp industries announced for investments of over 1.5 billion euros (Papernet.se, 2016a). Swedish Södra Cell is expanding fast in cellulose and textile pulp production (increase of 500,000 tons annually) with large investment during the last decade – e.g. more than one billion euro in two of their large, kraft pulp mills in Mörrum and Värö in 2015-2016. In Finland the so far single, largest forest industry project ever, will be started in the end of 2016 – the Äänekoski pulp mill of Metsä Botnia, an investment of over one billion euro – that will produce 1,3 million tons of pulp annually of which parts, initially, will go to large-scale textile, biocomposites/plastic and biogas applications in already established industry agreements, stabilized letters of intent and new value chains (Papernet.se 2016b). In the EU, the negative turning point for graphic papers can be dated to 2006/2007. Since then the printing paper segment has been reduced by 5-10 percent annually in the EU with recent
years (2012-2014) being particularly weak in that sense (EuroGraph, 2013; Johanesson, 2011) comparing to the global decline of 3% per year of printing papers.

**Technology shifts of papermaking and of diverging pulping and biorefining**

The industrial transformation of carbon for other purposes than construction, energy and paper has a long history. In the late 19th century (forest based) biomass was a natural feedstock for – of the times – large scale chemistry. After WWII that position was lost to a large extent due to the growing availability of cheap oil which in addition was easier to use for refining processes. Before this emergence of the oil based economy the chemical pulp mill – i.e. in this period dominated by sulfite pulping technology (figure 1) – was an important biorefinery infrastructure (Novotny and Laestadius 2014). It was to paraphrase Thomas Hughes, competing to become a dominant design for production of biomass based fuels, materials and chemicals, i.e. a biorefinery platform. A biorefinery is a process that separates and refines biomass compounds, in particular hemicellulose, cellulose (carbohydrate sugars) and lignin.

Cellulose is the most abundant organic, renewable material on earth, with hemicellulose and lignin coming shortly after at second and third place, all three large compounds in woody biomass (compounds holding 20–50% of the tree). Furfural for nylon, cellulose textile pulps, cellophane film and bioethanol were produced at integrated sulfite plants in North America and Germany in the early 1900’s with annual production rates of 20,000 tons (Kamm et al 2010).

After being neglected for decades the biorefinery concept had a revival a decade ago (1st International Biorefinery Workshop 2005; 1st-4th Nordic Biorefinery Conference 2009-2015; 1st International Conference on Biorefinery 2007; Larsson 2006; Kamm et al 2010). A biorefinery potentially has the advantage of producing more classes of products than petroleum refineries can, thus offering more opportunities for product development and diversification. In fact, products made of woody biomass are capable of replacing basically everything made of oil and many products made of cement, steel and aluminium (see Kamm et al 2010; Axegård, 2010; Lindström & Aulin, 2014, Novotny & Laestadius 2014).

The post-WWII decline of biorefinery related pulping processes was combined with a transformation of pulping technologies as well. The sulfite technology decreased in status substantially for three reasons. Firstly, the post war opening of markets and much lower oil prices made the sulfite process obsolete for production of transportation fuels (ethanol) and clothing (viscose versus cotton/polyester). Secondly, the kraft (sulfate) process progressively
replaced sulfite as the dominant chemical pulping technology as it processed all kinds of biomass, and produced stronger cellulose fibers for paper, thus contributing to the expansion of a global paperboard market that earlier had emerged in the US during the 1920s (Toivanen 2012). Thirdly, in the 1960s and 1970s, the sulfite technology was considered to be inferior also from an environmental viewpoint due to, by then, poor chemical recovery technologies. Hence, only a small amount of sulfite mills survived in countries with less demanding environmental regulations (Eastern Europe) or thanks to niche products in some applications (greaseproof paper; photo/film paper and dissolving pulps). A few, among them the Domsjö mill in Sweden (figure 1), managed to develop high class recovery technologies which contributed to its survival (Novotny and Laestadius, 2014).

The pulp and paper industry is characterized by continuous as well as semi-continuous processes. The latter involves batch processes of e.g. the digester and bleaching production processes. Continuous digesters are always used in modern large-scale (sulfate) pulp mills. The batch process encompasses the treating of bulk material in several batches through each step of the desired process. Processing of following batches need to wait until the current batch is completed. The advantage of the batch process is the possibility to carry out a greater variety of products or product characteristics at a production unit and also to obtain variation in volume. Batch systems, thus have a potential to produce pulp batches in the magnitude of 10s or 100s of tons a day rather than 1000s.

Large scale production of a single, uniform product, overall automation control of production processes and stable annual demand are some main advantages for the continuous process where input materials (chemicals, raw materials, etc.) are fed in the system at a constant rate, at a pre-set ratio and at the same time a constant extraction of outputs is made (quantity of main product, by-products, energy, etc.) (Kesner, 2001). In typical R&D environments such as laboratory and pilot scale the flexibility of batch processes is more commonly used. Biotechnology in production processes is seemingly easier to install with batch process technologies, e.g. enzymes in the bleaching steps of pulps. In addition, the diverging, multi-product and multi-purpose technologies of biorefining make batch processing attractive, at least for emerging biomaterials such as nanocellulose, carbon fiber from kraft lignin and enzymatic treatments. However, to scale up from a small lab test to tanks and vessels of ten or hundred cubic meters takes long processing time and may change product properties and reactants whatever process, batch or continuous, is used. This is basically what makes demonstration plants prior to market introduction a crucial step from bringing an invention in the lab to an upscaled innovation in process industries.
Figure 1. The sulfite process (technology of Domsjö, Borregaard among others) – a process for many bioproducts.

Figure 2. Kraft pulping process (with examples of biomass side-stream extractions)

Figure 2. Kraft pulping (dominant chemical pulping technology), a technology platform for diverging innovations and potential side-streams.
Box 2. (Thermo)mechanical pulping in the 1980s

The prodigious decade of mechanical pulping was the 1980s when printing paper segments demonstrated stable growth and a hunt for higher yields. The general low electricity prices – in North America and Scandinavia where TMP (thermo-mechanical pulping) is widespread – gave incentives to TMP to expand. TMP-technologies (and related newsprint production technologies) were upgraded significantly which made it possible to produce newsprint and magazine papers at low costs. Between 1980 and 1988 Canada expanded from 1.1 million tons of mechanical pulping to 6.55 million tons per year and in the same years Scandinavia expanded from 1.5 to 6 million tons per year (Biermann 1996).

On a general level the outcome of global upscaling of pulping processes and the fact they have been tightly coupled and integrated with the papermaking paradigm made the PPIs even more inflexible and fixed to the paper mill layout during the period of rather stable, global paper consumption growth between 1950s to early 2000s. PPI is today the largest biomass industry on earth with a global production capacity of almost half a billion tons per year, often located in plants with annual capacity of 200,000-1,200,000 tons. In fact, many incumbents still seem stuck into a “big is beautiful” paradigm. The cutting edge production units have since the 1950s doubled in production capacity every decade. Between 1950 and 1990 Swedish pulp and paper mills were on average among the most scale intensive. During this period Swedish mills on average went from roughly 15,000 to 225,000 tons per year (Skogsstatistiska Årsboken 1951; 1991). From 1990 this duplication of size in production units was overtaken by investments in South America, particularly Brazil, where new pulp mills reach 1,500,000 tons in annual capacity by year 2015. The contextualization in this section of diverging pulping technologies and the papermaking outcomes is important in order to understand the PPI related innovations and transformation pressures – as opportunities and necessities according to the Dahménian toolbox.

The nature of diverging innovations

In the previous section the technology shifts of mainstream pulping and diverging biorefinery processes were outlined. In this section I will explain the nature of diverging innovation of non-assembled products which I argue have profound policy and management implications in the emerging development block of biorefining and pulping.

This article has up till now been centred around the concepts of Dahmén and on historical and more recent technological shifts of pulp and paper industries. The PPI is in physical terms one of the largest, most capital-and resource intensive industries on earth, together with petro-chemical sector one of the industrial dinosaurs physically speaking – e.g. modern state-of-the-art plants producing more than one million tons per year. It is an industry run by business logics such as economies of scale, incremental innovations, high requirements regarding scale/capacity and stability of infrastructure – e.g.
transport systems, electricity and raw material supply – as well as run ability of the production system – i.e. coordinated, continuous, “perpetual” production processes with few breaks (Laestadius 1998; Novotny and Laestadius, 2014).

Another characteristic of non-assembled products is the high requirement on uniform product features and material composition. The material exactitude – the “recipe” of chemicals, molecules, etc., of viscose, steel, gasoline, or any non-assembled material produced with specific compositions, flows and temperatures in the production chain that may alter a successful pilot run of a new product when upscaling of technologies result in longer production chains and larger subsystems (pipes, tanks, vessels, etc.). This, in turn, increase the risk to alter production environment and, thus, product characteristics. As mentioned earlier this implies an increasing need for demonstrators to guarantee product characteristics that were found out in the lab also in the industrial plant, i.e. to bridge the product development gap from lab idea to market introduction. This is assumed as a result of demonstrators' abilities of running test batches in industrial environments which is more or less a requirement for process industries regardless which process, batch or continuous, is used (Joelsson, 2013; Storm, Lager & Samuelsson, 2013).

The production systems in process industries share similar characteristics (e.g. continuous, semi-continuous and batch material flow processes) which contrast radically from other manufacturing industries. The fact that production systems/processes are closely connected to product properties and specifications of manufactured products also affects innovation activities in a process industry such as pulping. Development work is here mostly managed in a laboratory/demonstration plant environment far from design studios of prototyping. In process industries, such as PPI and chemical industries, prototyping is not a significant development stage between idea and introduction stages, but instead replaced by test runs in pilot labs, demonstrators or full production industrialized test batches in order to verify process conditions. The often very long production chains of non-assembled products also make process and product innovation tightly coupled – e.g. glass-, paper-, and steelmaking (Novotny & Laestadius 2014; Storm, Lager & Samuelsson 2013).

The specific nature of diverging innovations introduced here has not been coined nor thoroughly explained by academia on technological transformations at meso-levels (industry/technological system) nor at management/firm levels. On the contrary, Utterback (1994) demonstrated how process industries developed by getting more and more integrated as long as technology progress proceeded. Although Utterback did not use the term this process integration – exemplified by integration of the plate glassmaking processes from several production steps to one continuous, automated
process – seemed to a large extent to form a path dependent track of “process industrialization” with high productivity gains and lower unit costs for each integrated step (Utterback, 1994, pp 106-116).

Linton and Walsh (2008) instead suggest that “product and process innovations occur simultaneously in material and process-based industries” which have implications for innovation management. In addition to their discourse, and quite differently to Utterback, it may be argued that process industries potentially can go “in the other direction” than the process integration and “process industrialization” of Utterback does. Instead, they may disintegrate and diverge. In other words, it would mean that process industries such as PPI may enter trajectories of diverging innovations with several side-stream or by-product opportunities, a “cascade alike”, product-tree characteristic similar to the material transformation flow of their production systems. In that case a diverging innovation opens up the production and innovation chain on at least one point – breaking the interdependence between pulp and papermaking chains – and putting the chemical pulp mill in focus. From there you may get a diversification towards different industries, formation of new business models and new competitors as seen previously in the biorefinery and Lignoboost cases in figure 1 and 2. This is what Dahmén probably would have labelled a transformation of a whole development block.

Also other, and today economically larger, process industries may face similar disintegration, such as oil, petroleum, and chemical industries. They may face the mirror image of PPI transformation competing for feedstock and markets. For example, oil and petro-chemical industries may soon have to replace fossils – petrochemical companies like Swedish Preem and Finnish Neste Oil partly already do – and instead put renewable, raw material feedstock in their production systems and try to modify logistics, chemical and process technologies, etc., along the chain in order to obtain the precise mix of molecules and bioproducts, now based on processed biomass. PPI, on the other hand, has renewable raw/input materials, infrastructure and large scale technologies in place for producing almost any chemical and material based on biomass. Exploit and move forward in the value chain is basically the next step for PPI endowed with chemical pulp mills. However, as long as oil prices, shale oil and gas extraction costs – as well as CO2 taxation – stay at the levels of recent years, biofuels and platform chemicals (e.g. biomethanol and other biobased hydrocarbons) remain hard to upscale. Biomass based trajectories – particularly so in regions with slow growing wood species like in the Northern hemisphere – will continue in the direction towards new biomaterials such as wood construction composites, lyocell, nanocellulose, lignin based products, etc. that are not necessarily competing directly with oil/petro-chemicals, such as resins, cellulose or lignin based packaging materials were regulation and consumers are more consistent (e.g. disallowance of plastic bags, regulation
encouraging recycling, wood construction, and consumer segments such as clothing).

In the case of material-based process industries, product features are strongly coupled to structure and the structure is strongly coupled to the processes — i.e. products are very seldom assembled (Linton and Walsh, 2008), and when assembled only downstream in converting/end consumer chains (converting consumer packaging processes). The number of co-evolved changes to the product and process will be greater than in assembly manufactured industries, since product innovation will necessitate a cascade of changes along the continuous production process.

Demonstration plants are in most cases crucial in large-scale process industries with high requirements on robustness regarding production process, transports and uniform product characteristics under certain conditions (e.g. chemical compounds or metals with high claims for uniform product characteristics along the production and logistic chain, etc.). Firstly because of the capital intensive and inflexible plant layout, secondly due to tightly coupled process and product innovations (non-assembled products), and thirdly as a result of the nature of diverging innovations which potentially may result in cascade alike, potentially “multiple purpose”, enabling technological innovations. Here, the examples of cascade-alike or diverging product tree of sulphite or kraft pulping (see figure 1 or 2) and other platform chemicals/biomaterials (see figure 4) have a myriad of applications which are not necessarily foreseen by the innovator and/or firm.

In demonstration plants the whole production chain with their extensively non-assembled products may be tested industrially at large scale in order to be able to replace large-scale fossil based products (mainly process industries of oil and petrol-based chemicals). This characteristic is even more augmented by the fact that production plants have high requirements on production availability and runnability (robustness) with huge costs in case of production breaks, a trend that has been reinforced in recent decades of intense competition (Abdulmalek 2006; Novotny 2007).

**Creative destruction and new development blocks**

During the last decade large corporations with integrated TMP and paper mills located in the Northern hemisphere, and positioned in declining graphic paper markets, are literally dying in the chimney fire. Printing and graphic paper demand has been halved in Western society in the last decade turning long production chains of papermaking processes obsolete (Hylander 2009; Johannesson 2011; Eurograph 2013). Niche players and incumbents of pulping for new applications and bioproducts displacing fossil based products in packaging, textile, chemical and construction industries are growing in the 2000s, but still in rather few segments and from quite low levels. Examples
are found in wood based textiles (lyocell and viscose in particular), wood composites and bioplastics, biofuels (blends in established biofuels) and a few platform biochemicals (e.g. furfural, lignosulfonates, and in general bioethanol for chemical applications, etc.).

First movers with biorefinery products based on the sulfite process – Borregaard, Lenzing and Domsjö – are now followed on larger scale by agents with other separating/fractionation technologies, which are the prerequisite for diverging innovations. An example of this is the mentioned Lignoboost technology which extracts the lignin – the second most abundant organic compound on earth – from the black liquor via CO₂ into a lignin chemical compound. Black liquor processed with Lignoboost technology at an average mill of a capacity of 500,000 tons per year would extract almost 100,000 tons of lignin of very high purity levels per year. Today two companies have implemented the technology (Stora Enso at the Finnish Sunila plant and Domtar in one of its North American plants which reach an annual capacity of 50,000 tons). In Sweden the potential is 1.6 million tons of pure lignin per year, and in the world almost 14 million tpy from unintegrated pulp mills and 25 million added with today’s integrated paper-pulp mills (Backlund & Nordström, 2014).

This new process technology opens up completely new, diverging “product-tree” families of lignin-based bioenergy products – heating value similar to carbon which reduces carbon footprint and corrosion in power plants based on coal – – and high value biochemicals – lignin phenols turned into carbon fiber, adhesives, foam additives or other high value chemicals with market prices of thousands of dollars/ton. In turn, it opens up new technological trajectories for special treatment of the other two main wood compounds – consisting of 70-75% carbohydrates, i.e. cellulose and hemicelluloses which basically can replace any product in chemical industries and fiber-wise even steel/metal applications (i.e. nanocellulose applications). As long as oil prices, shale oil, gas extraction costs and CO₂ taxation remain at the levels of recent years, biofuels and most biochemicals remain hard to upscale. Biomass based development – particularly so in regions with slow growing biomass species like in the Northern hemisphere – will continue in the direction towards new biomaterials such as lyocell, nanocellulose, lignin products that are not necessarily competing directly with oil/petro-chemicals. The dismantling of the “converging” and integrated process industry (Utterback, 1994) based on the gigantic capital intensive system results in an opportunity – a process of creative destruction (Dahmén 1970; Schumpeter 1981) – to make use of the diverging bioproducts of the pulp mill – the most large-scale infrastructure for biorefinery and separation technologies of today – , see figure 2 and in appendix figure 3 (Axegård, 2010; Kamm et al 2010). The diverging innovations of biomass have great potential of industrial synergy of sectors (energy, chemical, forest, automotive, agricultural, construction, biotech, etc.)
and as a consequence of strong transformation pressures mentioned in section 3 the development bloc of biorefining industries is opening up the closed process integration innovation à la Utterback at several points of the PPI chain (figures 1, 2, 4 and the long production chains of papermaking partly obsolete).

These are not simple signs of technology push driven innovations of a desperate industry trying to find new products and markets for their excess raw material base of wood due to papermaking decline. The macro-trends of societal and political pressures to reduce carbon emissions and to innovate environmentally sustainable bioproducts are real demands, although not always easily articulated and translated to the industrial realities of firms at the micro level. Subsidiary policy tools may contribute to a picture of regulatory contradictions and non-articulated policy tools (i.e. a Dahménian bottleneck for finishing a development block). Examples are found in Swedish energy policy that in some cases in the last decade have made it more profitable to burn primary round wood in bio-boilers for bioenergy instead of firstly, producing high value and longer lasting bioproducts (e.g. wood based cellulose textiles) which store carbon, and secondly, use biomass waste from processed and recycled cellulose fibers (e.g. worn out, recycled wood textile clothing) for energy.

Mainly societal pressures expressed by green consumer markets such as food, health, clothing and housing, however, now pave the way for the case of biorefining development block. Numerous successful examples are found in wood based textile. Lyocell and other cellulose textiles based on the chemical pulping technologies of Lenzing, Domsjö, Södra, among others, where new entrants (e.g. clothing companies only using organic material, or non-polymer based textiles) have filled the gaps and bottlenecks of the crucial distribution steps in the long chain of vertical linkages, described in section 2.1. The ongoing upscaling of the diverging innovation Lignoboost that extract and develop lignin based bioproducts as well as nanocellulose technologies stronger and lighter than Kevlar (thin bullet proof synthetic fiber) are two other examples of emerging segments that may find applications, primarily in process industries such as steel, petro-chemical (plastics and packaging in primis), automotive, and construction industries.

In the long run, diverging innovations – e.g. Lignoboost or nanocellulose – may also be disruptive for established pulping technologies. Examples of difficulties in managing disruptive innovations are numerous in history as the Schumpeterian idea of *creative destruction* demonstrates. Schumpeter himself stated that it was not the stage-coach companies that initiated the transition to the new railway construction business in the 19th century. The incumbents were inclined to exploit incremental innovations and external, new entrants the disruptive innovations initially considered too small and niched by incumbents. Such transitions might turn incumbents’ existing
resources out of date when new industries are emerging to respond to new, not always articulated demands. A process industry outside the graphic paper industry which faced an even more sudden decline is the example of large integrated steel mills that struggled to respond to the mini-mills in the 1980s and 1990s (Christensen, 2000; Tushman & Anderson, 1997; Schumpeter 1981).

The chemical pulp biorefinery – some possible products from wood compounds

Figure 3. Some examples of diverging innovations and industries from the largest organic, renewable wood compounds of the world: cellulose, hemicellulose and lignin compounds. Source: Axegård (2010), Kamm et al (2010)

Concluding remarks and implications

Academic implications

The sequential and structural mechanisms of industrial and technological transformation and its systemic nature have been possible to grasp with the use of the Dahménian toolbox. In this article this is expressed by a biorefining development block around which complex structural tensions and the industrial synergies are illustrated in times of “creative destruction” at different levels. Dahmén’s explanatory toolbox has resemblances with the structural path from Schumpeterian innovation theory. Although Schumpeter focused more on the innovation of the entrepreneur as the driver within an industry, Dahmén added the term development block which enables to identify transformation specifically across industry borders (linkages and tensions of demand pull across industries and technology push within an industry). They both shared the view on “creative destruction”, the transformation as a historical process of entrepreneurial activities, their new
products, market structures and the possible crowding out of old technologies, products and markets, etc. This structural analysis of business formation is most probably a heritage from Johan Åkerman, the supervisor of Dahmén (Laestadius, 2016).

The phenomenon of diverging innovation can be looked upon as both an empirical contribution: the breaking up of a closed integrated process industry towards something new, potentially several new or integrative industries, and as a theoretical comment on the model for non-assembled products in Utterback (1994). In the former, particularly in sectors where industrialist and managers are capable of foreseeing – ex ante – whole blocks of biomass as fundamental raw/input material and propelling linkages forward to refined bioproducts. Already there are signs of these transformations of diverging innovations, particularly in textile, construction, chemical industries, etc., that are breaking the process integration and industrialization of tightly coupled papermaking chains and also opening up petrochemical chains searching for renewable feedstock and “green” molecules. Instead of the process integration and “process industrialization” of Utterback’s innovation theory, these chains may disintegrate and diverge. This phenomenon is striking in the case of pulping and biorefining where agents are breaking the chains of papermaking. Diverging innovations in process industries could – particularly in times of complex transformations and structural tensions at different levels – take, and are taking in the case of biomass, on trajectories with several side-stream or by-product opportunities, a product-tree characteristic with analogies to the material flow of their own production systems. As seen in the PPI case diverging innovations, particularly in times of strong structural tensions, may open up the production and innovation chain on at least one point (e.g. recent capital spending of hundreds of million euros in chemical pulp mills in Scandinavia and adjacent separating and extraction technologies, such as Lignoboost). From there a diversification towards different industries from a wood-based perspective – and synergies with several different industries from a process industry perspective due to increased use of biomass raw materials and technologies from a multi-sectorial perspective. In addition, formation of new business models and new competitors as seen previously in the pulping and biorefinery case would be possible.

Innovation literature on industrial change may describe organisational consequences of disruptive transformation for incumbents. However, since mainstream innovation theory with very few exceptions has a focus on assembled products they at best only describe repercussions of transformation, but do not explain the intrinsic nature of innovation and transformation in process industries. Process industries such as pulping, steel, copper, chemical industries all rely on separating technologies with uniform product characteristics and composition (e.g. recipes of chemicals, molecules,
etc.). Technological transformation in these industries has overall the characteristic of potentially diverging innovations, i.e. intrinsically connected to the biorefinery technologies which per se may be defined as separation technologies of biomass compounds. As demonstrated – see figure 1, 2, 3 – the diverging innovations of biomass compounds may replace or integrate with other material-based industries (steel, chopper, construction, chemical, textile, food industries, etc.).

The toolbox of Dahmén (1970 and 1988) particularly outlined the complementary, vertical linkages across industry borders and technologies (identification of development bloc) in his explanation of industrial and technical transformation. Combined with the concept of diverging innovations in process industries which also describe vertical linkages as well as the potential for synergies and displacement of specific set of non-assembled products at the meso-level, a toolbox for industrial transformation of process industries is outlined. It is particularly useful in a period characterised by structural tensions at different levels – e.g. breaking up of an integrated process industry due to “creation of multimedia and destruction of a large part of integrated papermaking industry”, new biomass separation technologies and climate change mitigation – tensions that alter industry conditions and industry borders.

**Managerial and policy implications**

A consideration for management/policy implications is that market and technology shifts in process industries are rather about coupled process and product innovations where the outcomes are potentially diverging (and disruptive) under certain industry conditions. As an illustration, the Lignoboost technology invented by Swedish R&D organisation STFI in partnership with Chalmers university took more than ten years and several, upgraded pilot/demo-plants to commercialize. The nature of technological transformation in process industries at micro-levels is complex and seldom analysed at short-time basis. The case of Lignoboost, one of few successful and richly funded demonstrators, is striking. It was to a large extent supported by public funding for ten years as an “energy efficiency and recycling program” (the KAM-project funded between 1996-2002) and to offset bottlenecks in the recovery boiler (Axegård, 2010; Backlund & Nordström, 2014; Novotny, 2007). Two decades after, it is better described as a major diverging innovation and – at least potentially – a disruptive technology contributing to industry synergies and an emerging biomaterial/biorefining development block.

The diverging innovations of non-assembled products contain a different mind-set compared to innovation theories in general and specifically to those based on high-tech assembled manufacturing or ICT industries (cf. Utterback...
1994), which converge instead of diverge. Steel, chemical, copper, mineral and the pulping, biorefining case exemplified, are very much about separation technologies with a wide range of diverging product tree, cascade alike innovations, the opposite of innovations in assembled manufacturing. In the medium/long run this opens up for completely new development blocks of biorefining industries which can replace – and in the cases of wood-based clothing (lyocell and viscose) and construction (wooden skyscrapers) already are replacing – significant parts of today’s polluting and material intensive petro-chemical, construction, automotive and, clothing industries. In fact, the growth potential and projected market size for emerging bioproducts are often much greater than for traditional forest products combined (such as paper pulp, lumber and newsprint). Examples of annual growth rates of emerging bioproducts globally during a medium range period (2009-2015) are biochemicals (5.3%), wood fibre composites (10%), bioplastic and plastic resins (23.7%), platform chemicals (biomass based chemicals to make other chemicals, (12.6%), where biochemicals and wood fibre composites in 2015 entailed a global market potential of 100 billion USD annually (Natural Resource Canada, 2016). For wood-based textiles like lyocell, tencel, and viscose, growth rates have been around 7-8% the last decade. These wood-based textiles already have over 7% of the global textile market share and is expected to increase its share considerably (Svensk Papperstidning, 2015).

The nature of diverging innovation – if disruptive or incremental may be hard to judge depending on e.g. transformation pressures and structural tensions over time – in this kind of process industry urges policy tools both on micro- and meso-levels. On micro-levels different policies are needed for more favourable conditions in setting up development work in pilots, demonstrators and production plants, i.e. industrial environment for test runs of potential “products/processes” rather than policy tools for prototyping in design studios in order to lower the barriers to entry for innovators that seldom have access to biomass feedstock, R&D infrastructure and venture capital. Another policy consideration on a micro-level is that PPI companies which have changed strategies and market directions towards biorefining and biotechnology in Sweden, Finland and Canada have implemented technologies in production plants with batch digesters (Domsjö, Södra, Tembec, Domtar) or batch pulp bleaching steps (Finnish PPI). Batch processes hold the advantage of introducing more flexible and less costly demonstration plants in early stages of testing new products on new markets (initial niche markets) or to replace fossil based products with different requirements – a molecular composition of a biochemical instead of tensile strength of the paper machine web. This means that new functionalities and products require new demands on organisations and here the batch process at least technologically is more flexible than a continuous process for introducing niche products on new markets.
Addressed meso-level policies that encourage diverging innovations are more generic. In this case, policy tools for capability building of biorefining technology platforms that are hard for industrial organizations to create themselves; in general, for material technologies and process industries in the creation of new development blocks. This may include potential infrastructures for production/distribution chains and their main agents who are supposed to think in blocks *ex ante* or that are able to see potential blocks and vertical linkages between biomass-based industries and other large-scale industries moving away from fossils. Innovation policies often contain a misconception to over-simplify entrepreneurial events by conceiving them as individualistic practitioners rather than important agents in function of a sound entrepreneurial “eco-system” (Stam, 2015)

Complications with IPR (Intellectual property rights) causing patent thickets in this transformation of the up-scaling phases of R&D projects, e.g. after promising pilot-scale attempts may benefit from “ambidextrous” organizations managing IPR pools of patenting members if bottlenecks emerge as a result of patent thickets.

**Suggested further readings**


Methodological considerations

The data collection approach is qualitative, based on case studies, observations from plant visits and semi-structured interviews – predominantly interviews that also made the empirical basis for other articles published back in 2013-2014 (see Novotny & Nuur 2013; Novotny & Laestadius 2014). Observations and insights build on industry statistics, on approximately 30 visits to PPI mills, and to conferences were completed between 2005-2009 as editor of a PPI magazine and during 2010-2015 as Ph.D. candidate (including informal discussions at mills and conferences). The interviews/visits facilitated technological insights on PPI technologies and industry segments. Managers as well as R&D staff were asked about new product fields/technologies, industrial investments (i.e. new equipment, pilot machines, demonstrators, new products, etc.), R&D, etc. There is also a growing research literature in the biorefinery field of engineering which has been utilized in the analysis (see eg. Kamm et al 2010; Axegård 2010; Backlund & Nordström 2014). Several interpretations have been made in order to capture the transformation of pulp and paper industries in the Northern hemisphere (Canada, USA, Sweden and Finland in primis). They are analysed in terms of general similarities and differences (e.g. pulping technologies) from which management implications subsequently could be drawn. Qualitative approaches have the benefit of a limited set of observations that may capture a large number of aspects with regards to the studied case, for example, to get into the black box of technology, diverging innovations of pulping (i.e. biorefining), and causal mechanisms behind transformation processes (Dahmén 1950/70; Yin 1994).

References


Papernet.se (2016a) Investeringsboom väntar i massaindustrin. 22/5-16 [accessed on-line 15/6-16: http://papernet.se/alla/investeringsboom-vantar-massaindustrin/]

Papernet.se (2016b) Textil i nya Äänekoskifabriken. 29/3-16. [accessed on-line 20/4-16: http://papernet.se/alla/nyheter/textil-i-nya-aane-koskifabriken/]


Money manifested in cash is a social innovation that has had tremendous impact on the development of mankind but we now see that cash is starting to be questioned in some markets, and I aim to develop a better understanding of this process.

Drawing on theories of change in socio-technical systems, I provide an illustrative example of a system – cash-based payments in Sweden – that is challenged by developments in niches and that therefore may become radically changed via disruptive innovation and potentially be replaced by new services. When doing this, I answer two questions. First, which are the most important factors influencing the use of cash, and how does each of them affect the use of cash? Second, which implications for policy-makers and managers can be drawn? I thus aim to provide a deeper understanding of the process potentially moving Sweden towards a cashless society.
that a central actor which today is the Riksbank, the Swedish central bank, guarantees the value of bills and coins denominated in Swedish Crowns (SEK).

Bills and coins are unilateral contracts where the central bank – and in the end the government in a nation – guarantees a quid pro quo exchange of each bill and coin to an identical value in an electronic account. The guarantee that was issued on clay coins by King Ammi-ditana is basically the same as the Riksbank’s guarantee today. This is particularly transparent when studying banknotes from England’s central bank – the Bank of England – on which it is written that: “I promise to pay the bearer on demand the sum of 20 pounds”, which is the guaranteed by the Bank of England and supported by the Queen of England. The English banknotes were in fact called promissory notes, i.e. noted promises, from the beginning which manifests the unilateral promise.

But this promise – as well as all other promises – must of course be credible to have some value for the parties involved. The credibility of the promise that a bill can be exchanged for something else of equal value is therefore critical to how the public will view and use cash. Today this credibility is fundamentally based on the relationship between a nation’s actual underlying assets (for example. GDP) and expectations of future economic development (e.g. the capitalization of the stock exchange and/or total outstanding loans in the banking system). And the value of cash as a useful payment service is also related to the inter-operability of cash, i.e. how many payers that want to use cash and how many payment receivers that accept cash as a payment service. The fewer payers that wants to use cash, the less valuable will it be for a payment receiver to accept cash, and vice versa. The lower the inter-operability of cash, the lower the efficiency of the payment service and also the lower is the value of the service (Economides, 1996; Hagiu & Wright, 2015).

Today there is a growing concern whether cash-based payment services are efficient – at least in the Scandinavian countries. The costs of different payment services in our societies have led to an intensified debate on the need for a reduction of cash-based mass-payments in Sweden. Macro-economic studies show that the economic benefits for a society if cash is replaced by electronic payments may be around or even more than 0.3 percent of GDP annually (Segendorf & Jansson, 201251; Danmarks Nationalbank, 2011). Other studies show similar results (Bergman et al., 2007; Humphrey et al., 2006). In addition to this, there is growth of service innovations that have the potential to substitute cash-based payments (Arvidsson, 2013).

Given the importance of the payment system for the entire society and the potential cost saving potential identified in economic studies, I analyse how

51 The study by Segendorf & Jansson (2012), however have been criticized for underestimating the costs of card payments by Hortlund & Svensson (2013).
processes of industrial dynamics and its driving forces can explain the development of the landscape for cash based payments, and draw conclusions for policy-making. The article has two research questions:

1. Which are the most important factors influencing the use of cash, and how does each of them affect the use of cash?
2. Which implications for policy-makers and managers can be drawn from our study?

**Cash-based payment services**

The Riksbank, the central bank of Sweden, states that the: “the financial system has three main roles: to convert savings into funding, to manage risks and to make it possible for payments to be made efficiently” (Sveriges Riksbank, 2012:6; cf. Willesson, 2007:12). My study is limited to the third role, i.e. efficient and secure payments, which is critical for the entire society since the: “smooth performance of financial transactions is important if the economy as a whole is to function efficiently” (Sveriges Riksbank, 2012:9). And more particularly, I will focus on one payment service, namely cash.

The Riksbank divides payment services into different categories based on two dimensions: First, if the value behind the payment is direct, i.e. that the value is stored in the payment instrument in itself, or if the value resides in a separate account to which the payment instrument is connected. These categories are the horizontal ones in table 1 below, namely account-based or direct. The second dimension relates to the technology behind the service, i.e. whether it is electronic or paper-based (see table 1).

<table>
<thead>
<tr>
<th></th>
<th>Account-based</th>
<th>Direct</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Electronic</strong></td>
<td>Credit transfers, Direct debit, Transfers, Debit cards, Charge cards, Credit Cards, E-money and Mobile payments</td>
<td>Prepaid cards, E-money</td>
</tr>
<tr>
<td><strong>Paper-based</strong></td>
<td>Paper-based credit transfers, Money orders, Checks(^{52})</td>
<td>Cash</td>
</tr>
</tbody>
</table>

*Table 1. Types of retail payment services in Sweden (Source: Sveriges Riksbank, 2012:120)*

My study focuses on direct, paper-based payments in Sweden, i.e. cash payments made in Swedish crowns (SEK), which constitute a subset of retail payments. Cheques are used extremely seldom in Sweden.

\(^{52}\) Cheques are used extremely seldom in Sweden.
payments, i.e. payments that occur in large numbers but where the amount per transaction is relatively small (Sveriges Riksbank, 2012: 116). In practice, factors related to other types of payment services will also be part of the analysis as long as they have bearing on the industrial dynamics related to cash. The other services can for instance be – more or less perfect – substitutes to cash payments and therefore be critical for the development of cash-based payments. In addition, there is regulation and actors that are connected not only to cash but also to the other categories of payments.

A practically oriented definition of the payment system can be formulated as a system built on trust for both individual companies and their systems that handles payments including actors both within the transparent regulatory framework (“open loop”) as well as actors with their own regulatory framework (the ‘closed loop’)53. Another important distinction is that the study mainly covers proximity payments in the retail sector, i.e. payments of small amounts that occur many times.

A broad definition of the technological innovation system for cash based payment would include banks (and related organizations such as trade unions, Bankgirot, banks, Bankernas Depå, Bankernas Automatbolag, etc.), card issuers, card acquirers, card companies (e.g. VISA, Mastercard, American Express etc.), providers of payment services both for consumers (e.g. Pay Pal, Izettle, etc.) and for companies (eg. Payex, Payair, Seamless, etc.), suppliers of clearing services (Automated Clearing Houses), mobile network operators offering payment services, “money transfer”-companies (eg. Western Union etc.), companies supplying cash handling services (e.g. Kontanten, cash-in-transit companies, etc.), new entrants or “invadors” (Ramirez & Wallin, 2000) (e.g. Google), government agencies (e.g. Sveriges Riksbank, the Swedish financial supervisory authority, the tax authority, etc.) as well as actors and sectors that directly or indirectly uses the technological and regulatory platform for payments in their own business models (e.g. merchants like IKEA and ICA, Swedish Commerce and trade unions, public transportation companies like SL, UL and SJ, hotels, the travel and entertainment industry, restaurants, etc.). In addition, it would include consumers. A more confined

53 The open loop is the part of the payment system that is governed by Sveriges Riksbank. This is what we generally talk about when discussing payments. The closed loop includes transactions that are handled within a closed system, and are not directly transparent in the official system. Examples of closed loop payments are bonus cards given to loyal customers and used as payments of goods and/or services, or coupons that only can be exchanged with goods and/or services. This also includes local currencies like Llewes pound in the UK and Dalern in Sweden. Closed loops are governed by the issuers of such payment schemes and not directly controlled by the Sveriges Riksbank.
definition of the cash based system for retail payments – which is the one I will use in this study\textsuperscript{54} – focus on the open loop transactions.

Statistical data on the use of cash-based payments in Sweden provides an understanding of the state of development of the TIS for cash-based payments. The proportion of cash payments in the economy is somewhat difficult to measure because there are no statistics on, for example, person-to-person payments. A generally accepted indicator of the use of cash is therefore the value of outstanding notes and coins in relation to GDP. This figure for Sweden was around 10 percent in 1950 but has dropped to 2.6 per cent in 2012 (Sveriges Riksbank, 2012) and even below 2 per cent in 2016. This is a result both from a relatively strong growth of GDP but also from an actual reduction of the nominal value of outstanding cash. The absolute value of outstanding cash in Sweden peaked at around 109 billion SEK in 2007 but has dropped to around 94 billion SEK in 2012 (Sveriges Riksbank, 2012), and even below 70 billion SEK in 2016 (Sveriges Riksbank, 2016). If comparing Sweden to other European countries, it is evident that the use of cash in Sweden is very low. At the end of 2010, the value of outstanding Euro-cash (notes and coins) was around 9.4 percent of the total GDP for the Euro region\textsuperscript{55}. In addition, interviews with merchants indicate that around 20-30 percent of the proximity payments in retail industries are done with cash and the rest is done with card payments.

\textit{Industrial dynamics related to cash-based payment services}

I frame my study in theories on innovation systems and innovation dynamics. Research on dynamics in innovation systems has concluded there are factors conserving a system as well as factors changing the system where dynamic processes are shaped by the interaction between these two types of factors. Such factors include sociology and institutional factors (Geels, 2004), technologies (Bergek et al, 2007), and technological- and business ideas (Arvidsson & Mannervik, 2009; Johnson & Jacobsson, 2001). The theories are built on frameworks where parallel processes either makes the current industrial system function in a more efficient manner in the short run or makes it change in a more radical way in the longer run (Schumpeter, 1934; March, 1991; Geels, 2004; Bergek et al, 2007; Arvidsson & Mannervik, 2009). These theoretical models suggest that the cash-based payment system will be affected by such parallel forces where some lead to increased efficiency and others lead to radical innovation (Henderson & Clark, 1990), but also that

\textsuperscript{54} The open loop definition will restrict the analysis to the traditional industry supplying payment services based on accreditation by the Swedish Financial Supervisory Authority (Finansinspektionen) and governed by Sveriges Riksbank.

\textsuperscript{55} The value of outstanding notes and coins in the Euro zone (EU17) was around 862 billion Euro in 2010 and GDP for the same region was 9 204 billion Euro (sources: Eurostat and European Central Bank).
interaction between these two sets of forces in itself will affect industrial
dynamics and ultimately the use of cash-based payment services.

My analysis of the payment system is based on the definition of a socio-
technical regime by Geels (2004: 900) who: “... define ST-systems in a
somewhat abstract, functional sense as the linkages between elements
necessary to fulfil societal functions (e.g. transport, communication,
nutrition). As technology is a crucial element in modern societies to fulfil
those functions, it makes sense to distinguish the production, distribution
and use of technologies as sub-functions. To fulfil these sub-functions, the
necessary elements can be characterised as resources. ST-systems thus
consist of artefacts, knowledge, capital, labour, cultural meaning, etc.”. I aim
to identify such elements and how they affect change of the STS for cash-based
payments, and in this manner outline the socio-technical regime, i.e. “... the
‘deep structure’ or grammar of ST-systems, and are carried by the social
groups” (Geels, 2004: 905) for cash-based payments.

I consequently base my research on the definition of a system as a sectoral
innovation system by putting emphasis on: “the structure of the system in
terms of products, agents, knowledge and technologies and on its dynamics
and transformation. In broader terms, one could say that a sectoral system
is a collective emergent outcome of the interaction and co-evolution of its
various elements” (Malerba, 2002: 251). The system is driven by a collection
of organisations, people, competences and interests that collaborate and
compete in different constellations, which also may change over time. This
approach is in line with Moulaert & Sekia (2003) and Martin & Sunley (2003)
who ask for models of territorial innovation that addresses dynamics and
evolutionary dimensions of innovation processes.

My choice of this theoretical approach (i.e. Geels, 2004; Geels, 2002) is
motivated by the basic characteristics of the industry. Payment services in
general and especially cash-based payment services have characteristics such
as strong regulation and supervision by the state in the form of a policy regime,
particular technology regimes related to different types of payment services, a
clearly defined and important user and market regime both in terms of payees
and payers, a strong and important socio-cultural regime related to the view
of cash, as well as a – arguably less evident – science regime related to
research and development in the industry (see Figure 1).
Based on Geels framework, I sort critical factors influencing the industrial dynamics related to cash-based payments in three different categories: landscape developments, socio-technical regime and technological niches. The landscape developments are related to the entire payment system (i.e. all four categories in table 1 above), the socio-technical regime is more directly related to cash, and the technological niches are related to new technological factors that may lead to the creation of substitutes to cash.

**Industrial dynamics in cash-based payments in Sweden.**

To understand dynamics in the cash-based part of the payment system, I first need to define the critical concepts and actors. I start by identifying actors using a report on cash handling in Sweden provided by Sveriges Riksbank (2011 (a)) as the base for identifying actors in the industry. The report identifies the following key actors in the system for cash handling:

- **Banks** – whose main tasks relates to cash withdrawals and deposit from and to bank accounts via ATMs as well as manual services in bank offices.
- **Depot services** by Bankernas Depå AB (BDB) – whose main responsibilities are to collect cash from and deposit cash to the Sveriges Riksbank and to store cash in depots. This company is jointly owned by the major banks in Sweden and has outsourced the operational responsibility of these depots to cash handling companies. Banks, cash handling companies and providers of
automated cash handling services are in practice customers to this depot company when buying cash from it.

- **Providers of automated cash handling services** – whose main responsibility is to offer and operate ATMs and automated deposit machines. One company in this area is Bankernas Automatbolag AB (BAB) that is jointly owned by the largest banks, while another is Kontanten AB that is privately owned. The customers to these services include card acquirers, banks and merchants.

- **Cash handling service providers** – whose main responsibility is to operate cash depots, provide transportation and security services, operate automated cash handling machines, administer cash control services (counting, controlling authenticity and providing quality control) and supply cash to banks and merchants. The main companies in this area are Loomis and G4S Cash Solutions. There are also companies specializing in niches like collecting coins from parking meters such as Nokas Cashhandling and Skandinavisk Mynt Transport.

- **Receivers of cash payments like merchants and other payees** – whose main responsibility is to accept cash as a mean of payment for product and/or services and therefore also to administer cash in the period from when a payment is made until the cash is deposited in a bank account, which often is the same time as when the cash handling service providers receive the cash. Today merchants and other payees also provide cash distribution services similar to an ATM withdrawal in connection with a payment and therefore provide a form of a cash withdrawal service. It should be noted that while payees traditionally only saw cash a mean of payment they now also provide cash related services that they can build a business case upon.

- **Payers** – i.e. consumers and others who prefer using cash as a payment service in specific situations and at specific times.

In addition, the system includes producers of bills and coins, sub-suppliers of raw material, suppliers of machines and technology (e.g. ATMs), suppliers of services that enable ATM usage (such as card issuers, infrastructure providers and others), and other financial institutions. These actors play a less dominant role and will therefore not be discussed in detail.

But there are also important institutions governing cash-based payments. One important actor is Sveriges Riksbank. Its role in the payment system in Sweden is built on the premise that Sveriges Riksbank has responsibility for Swedish bills and coins (RBFS, 2009:3; Lagen 1988:1385 om Sveriges riksbank) which among other things includes the supply of coins and bills, the destruction of non-useable coins and bills as well as handling false money. To enable this, Sveriges Riksbank has the responsibility to supply cash that then is transported to depots and their representatives who in turn distributes this

---

56 There was a third large provider of cash handling services – Panaxia – but this company went bankrupt during the fall of 2012.
to banks and retail stores. Cash is then distributed and re-distributed continuously until they are non-usable and finally destroyed. Cash handling service providers play an important role in these processes. All in all, the cash-based system circulated 349 million bills with a value of 91 billion SEK and 1,9 billion coins with a value of 5,3 billion SEK in Sweden in the end of 2012 (www.riksbanken.se). In 2016 the value of outstanding cash in Sweden had decreased to below 70 billion SEK.

It is important to note that the amount of cash in the Swedish economy is not regulated by Sveriges Riksbank but instead determined by consumers’ actual demand 57. This decentralized and market-oriented system provides opportunities for banks and depots to exchange and trade cash by themselves. There are also opportunities for these actors to return excess cash to Sveriges Riksbank. There were 12 bank-owned cash depots in operation in Sweden in January 2010 (www.riksbanken.se). As stated initially, a result from this decentralized system is that bills and coins are distributed and re-distributed continuously in the society as economic transactions are initiated. In addition, there are numerous transactions between firms and consumers as well as between consumers. The easiness and swiftness with which cash stimulates transactions, trade and economic prosperity also has a flipside. Cash is an excellent tool for transactions that people want to keep secret.

In Sweden there is a tradition that consumers should not be charged for using cash which has led to an excess use of cash (Arvidsson, 2013). Guiborg and Segendorff (2007) show how banks’ net results from cash and checks are negative which, however, is compensated by an equally large but positive net result from card payment services. The overall net result from payments in banks is also strengthened by a positive result from credit transfers. Their study shows that banks cross-subsidize cash transactions with profits from other types of payment transactions, which lead to situation where consumers tend to use more cash than what is economically optimal.

Studies show there is a need to make the cash-based system more efficient (Segendorf & Jansson, 2012; Danmarks Nationalbank, 2011) but there are, at the same time, weak incentives for actors to seek innovation and development, which stifle entrepreneurial experimentation. Cash also has a strong formal legitimation in the sense that cash — by law58 — is legal tender. There are, at the same time, problems related to criminality and environmental issues that

57 Even if there at the same time is a public responsible to supply opportunities for all citizens in Sweden to be able to make payments which in many cases mean that cash-based payment services must be provided. This is controlled by county governors (see Länsstyrelserna, 2011).
have made actors doubt the usefulness of cash. Cash is also not always a good business case. Actors directly focused on technologies and services enabling cash-based payments generate resources and profits that enable investment in continued development while other actors – like banks – that supply cash services do not generate profit from these activities and therefore has little incentives to develop these activities. To be clear, many actors – like banks – would rather see an expansion of card-based payments – where they can make profits – than a continuation of cash-based payments.

All in all, my study identified a number of mechanisms that explain the use of cash in Sweden via the literature study and interviews which then were discussed in the workshop meetings and presented in Arvidsson (2013). Table 2 provides a list of these factors by outlining important factors influencing industrial dynamics in this sector but are also related to the main dimensions in Geels’ model. This provides an answer to my first research question, i.e. which are the most important factors influencing the use of cash, and how does each of them affect the use of cash?

<table>
<thead>
<tr>
<th>DRIVER / BARRIER</th>
<th>CONNECTION TO STS (GEELS, 2004)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Factors driving change toward a cashless society</strong></td>
<td></td>
</tr>
<tr>
<td>Technologies for and substitutes to cash (including characteristics)</td>
<td>Socio-technical system</td>
</tr>
<tr>
<td></td>
<td>Human actors, organizations, social groups</td>
</tr>
<tr>
<td>Communication and learning</td>
<td>Human actors, organizations, social groups</td>
</tr>
<tr>
<td>Digitalization / platforms</td>
<td>Socio-technical system</td>
</tr>
<tr>
<td>Demographics</td>
<td>Human actors, organizations, social groups</td>
</tr>
<tr>
<td>Crimes and policies to reduce crime related to cash</td>
<td>Human actors, organizations, social groups</td>
</tr>
<tr>
<td></td>
<td>Rules, institutions</td>
</tr>
<tr>
<td><strong>Factors acting as barriers to a cashless society</strong></td>
<td></td>
</tr>
<tr>
<td>Consumer behaviour (habits, demand)</td>
<td>Human actors, organizations, social groups</td>
</tr>
<tr>
<td>Business strategies related to payments</td>
<td>Human actors, organizations, social groups</td>
</tr>
<tr>
<td>Level playing field for payment services</td>
<td>Rules, institutions</td>
</tr>
<tr>
<td>Politicians and public debate on cash</td>
<td>Rules, institutions</td>
</tr>
</tbody>
</table>

Table 2. Drivers towards a cashless society and barriers to a cashless society in Sweden
The analysis of interviewees’ responses indicates that the strongest drivers toward a cashless society are “Technologies for and substitutes to cash”, “Communication and learning”, “Digitalization / platforms”, “Demographics” and “Crimes and policies to reduce crime related to cash”. The two absolutely strongest drivers for change towards a cashless society are “Technologies for and substitutes to cash” and “Communication and learning”.

The analysis of interviewees’ responses also indicates that the strongest barriers toward a cashless society are “Consumer behaviour (habits, demand)”, “Business strategies related to payments”, “Level playing field for payment services”, and “Politicians and public debate on cash”. Two of these – “Level playing field for payment services” and “Politicians and public debate on cash” – stand out as the two absolutely strongest barriers.

Geels (2004) identify three dimensions in a socio-technical regime – rules and institutions, socio-technical system, as well as human actors / organizations and social groups – which can be used to deepen the understanding of change in the cash-based payment system. Based on these results, I will illustrate how the most influential change factors (in accordance with Bergek et al (2007) and Arvidsson & Mannervik (2009)) for the cash-based payment system relate to the dimensions in Geels (2004: 902-903) model, i.e. socio-technical system; human actors, organizations and social groups; and rules, institutions.

I conclude that both companies and policy actors can play important roles in relation to the most important factors identified in the analysis. To start with, policy actors definitely can play an important role in relation to many of the factors. As for the ones driving toward a cashless society, policy actors like the department of finance and Sveriges Riksbank (in their roles as actors) could play a role in how to communicate the importance of how and which payment services are used in the Swedish society. The police, BRÅ 59 and the tax authorities have, of course, critical roles in how to prevent and handle crimes related to cash. In addition, universities – private or public – can play important roles in relation to both “Technologies for and substitutes to cash” and in relation to “Digitalization / platforms”.

As for the most important factors being barriers to a cashless society, rules and institutions are by default the key dimension. I conclude that politicians and public organizations (in the form of rules and institutions) play critical roles directly in relation to the barriers to a cashless society, and can play important roles (in the roles as actors) in relation to some of the drivers toward a cashless society.

59 A crime prevention agency.
Business actors, on the other hand, have a relatively stronger role in relation to the drivers toward a cashless society if compared to their role in relation to barriers to a cashless society. Companies play key roles in creating “Technologies for and substitutes to cash” as well as “Digitalization / platforms”. Both these processes can of course be done in collaboration with universities and researchers. Business actors can also play important roles in relation to influencing “Consumer behaviour (habits, demand)” via education and information as well as via incentives provided to consumers in the firms’ offerings and business models. In the end, however, the companies’ control on of the most decisive factors potentially preventing the formation of a cashless society, i.e. “Business strategies related to payments”. It is evident in the interviews as well as in the work done in the workshops that incumbents like banks that today basically have no revenues from cash-based payment services and runs a profitable business around electronic payments like card payments should have strong incentives to change their business model in the form of pricing cash-based services in relation to the costs of running these services. The media and consumers do not seem to allow this, however. There is a strong tradition in Sweden that cash-based payment services should not carry a price (Arvidsson, 2013) and this tradition is difficult to change. Paradoxically, the banks also seem to see a potential strategic threat in a situation where cash-based payment services are replaced by electronic payment services provided by others than banks, which makes the change of strategies and business models even more complicated. In sum, even if actors like banks may control one of the most important factors preventing the reduction of cash – “Business strategies related to payments” – they seem to have large difficulties in knowing to act in this matter.

Policy implications

My final contribution is to draw policy implications based on the results in my study and to answer my second research question, i.e. which implications for policy-makers and managers can be drawn from our study? I will use the results from the analysis above and additional data on implications from a cashless society to identify and outline important policy issues. This data will be used as indicators of relevance of the policy recommendations that the analysis provide. A general conclusion is that the policy recommendations in relation to the coming of a potential cashless society have bearing on politics since the most likely implications from a cashless society relates to reduced criminality and increased efficiency in mass payments.
Table 3. Identified implications from a cashless society

<table>
<thead>
<tr>
<th>Implication</th>
<th>Numbers</th>
</tr>
</thead>
<tbody>
<tr>
<td>Security costs would decrease; fewer robberies</td>
<td>7</td>
</tr>
<tr>
<td>Increased efficiency in mass payments (and fewer players)</td>
<td>5</td>
</tr>
<tr>
<td>Banks will have a new role</td>
<td>3</td>
</tr>
<tr>
<td>Real time processing, clearing &amp; settlement will come; Greater demand on consumers to select the best service; Risk of higher fees for payment services</td>
<td>2</td>
</tr>
<tr>
<td>Card payments will lose its dominance; Possibilities for MBT; A need to invest in POS terminals and other equipment; A need to educate employees handling payments as well as consumers; Higher tax revenues for the state; New value-constellations with actors supplying mobile payment services; A need to solve collaboration for infrastructure/interoperability while competing on (mobile) payment services; Increase of e-commerce which stimulates mobile payments; Importance of cash in a crisis situation; Brave politicians that drive change; Spot market for payment services (that can be decided on each payment situation); and New infrastructure for mobile payments</td>
<td>1</td>
</tr>
</tbody>
</table>

The first policy implication concerns how to develop a payment system that is robust and can withstand crisis in the form of power breakdown, hacker attacks and other types of problems that makes payments impossible, while at the same time providing better opportunities to prevent and minimize crimes related to payment services. This encompasses the creation of electronic payment services that can substitute cash-based payments in areas where these are strong today. In the end, this work must aim to ensure an infrastructure that realizes the possibility of making payments even in the case of electricity and/or Internet black-outs.

The second area is related to increased efficiency in mass payments. Here, the main issue so to ensure a level playing field for competitors in the field of payment services. This involves understanding competition and anti-trust issues also when new actors from other industries than the financial industry enters the scene. This should be governed by the notion that each payment service should bear its own costs and that these costs are taken by the users of each service. To achieve this, the work has to acknowledge two important issues. First, it is important to ensure transparency regarding the border between collaboration for inter-operability and competition for efficiency. Any substitute to cash must be interoperable in order to provide possibilities for economies of scale and scope but at the same time allow competition between providers of new types of payment services. Second, these ambitions...
must of course also aim to realize the objectives behind Payment Service Directives I and II as well as the Single Euro Payment Area to stimulate increased international and national competition within payment service industries in EU.

Increasing efficiency also relates to innovation and there is room to actively work to stimulate a growth of an innovation system for payments by actors like VINNOVA\textsuperscript{60} in Sweden. The immediate objective would be to stimulate innovation of electronic substitutes to cash in order to reduce the economic costs of cash-based payment but also to – synergistically – stimulate a growth of a Swedish cluster in the payment service industry. Such efforts could also be directed at solving the rising problems for people that rely on cash as their main, or perhaps only, possible way to make and receive a payment. Public procurement could, for instance, be used to direct innovative efforts toward solving the problem of how elderly, unbanked, people with disabilities and immigrants can make payments and receive money when cash disappears\textsuperscript{61}.

Initiatives to use public procurement to solve this problem is actually taking place in Sweden in 2016.

Another area is related to communication and learning. It is important to make sure consumers are prepared for a fundamental change in the payment system. This includes educating those who use a lot of cash today as well as educating consumers to become good customers in a changed business environment where we may see increased competition between providers of payment services as well as between types of payment services. And, where growth of different types of set-ups where some are direct transfers between accounts and others are based on credits, and the pricing structures include annual fees as well as fees per transaction will lead to a fragmented industry and a complex situation for consumers. The objective of education and information must be to create stronger customer power and ability to select best suppliers. This aims both to ensure a healthy competition in the industry via demanding and insightful customers as well as to minimize the risk that consumers make poor choices in terms of selecting inadequate payment service providers and type of payment service.

To sum up; the challenges connected to a potential reduction of cash-based payment services can be described as, first, a policy-oriented challenge to ensure a level playing field for competitors in the field of payment services. Second, another policy-oriented challenge to develop a payment system that is robust and can withstand crisis in the form of power breakdown, hacker attacks and other types of problems that makes electronic payments impossible. Last, a combined challenge for policy-makers and firms wanting

\textsuperscript{60} A governmental innovation agency

\textsuperscript{61} See: www.dn.se/debatt/repliker/stefan-ingves-har-en-gammaldags-syn-pan-appengar/
to sell payment services to make sure consumers are prepared for these fundamental changes in the payment system, and that there are services attractive for all types of consumers.

I also would like to emphasize that different actors in the system can and should do different steps to influence change in the cash-based payment system. First, the results illustrate the complexity of change due to the fact that both policy-makers and firms play important roles in the change processes, and that these may be complementary but in some cases also contradictory. Business actors have a stronger influence on factors driving change while policy actors have a stronger influence on factors inhibiting change. This is not a surprising finding but it means that policy-makers and business firms must – in some way – collaborate in order to avoid a situation where the system – or field – is stuck in inertia.

This implies that the overall change in the system can be understood by the tensions and interactions between factors driving change, which tend to be controlled by private actors, and factors acting as barriers to change, which tend to be controlled by public actors. This does not say that each of these two types of actors – private and public – only will be connected to drivers or to barriers. But there is a tendency that each type of actor has a stronger connection to drivers of change or barriers to change. The tension could potentially be understood as situations where public and private actor belong to different “thought worlds” (Janis, 1972) and therefore have problems to create synergistic collaboration for change in this system. I therefore believe the research on industrial dynamics in socio-technical systems could perform deeper analysis of how tensions between different elements and actors influence industrial dynamics. Such processes should aim at avoiding inertia (Arvidsson, 2014).

The socio-technical system around cash provides an illustrative example of a system – the payment system in Sweden – that is dynamically stable and undergoing radical change leading to a new landscape. We still do not know exactly how the new landscape will look like but it is most likely characterized by very few – or potentially even none – cash payments.

**On methodological considerations and the use of sources in this chapter**

I have collected data via a literature review, interviews and interactive meetings where participants discussed findings from the literature and interviews. The data is based on 17 interviews during the fall 2011 and spring 2012 with representatives from banks, retailers, service and infrastructure providers, card companies, trade unions, Sveriges Riksbank and other public
authorities attached to the payment system. All in all, there were six interviews with banks, three interviews with payment service providers, six interviews with merchants (including consumer retail stores and hotels), and one interview with one union. In addition, I also held three interactive meetings (or workshops) where each consisted of a half-day meeting with 6-13 participants. These representatives in these workshops included representatives of four banks, three payment service providers, Sveriges Riksbank, two different unions (one related to banking and one to merchants), six different merchants (including consumer retailing, hotels, parking, public transportation), and two representatives from a municipality in Sweden. All in all, the workshops included 24 external participants in addition to two researchers.

I must also acknowledge the delimitations of my discussion. A payment system is generally separated geographically by national borders, i.e. each nation tends to have their own currency and consequently issue their own bills and coins, which also is the case when I have studied cash-based payments in Sweden. A study of cash-based payments in Sweden is mainly focusing on factors affecting the Swedish payment system and Swedish money (SEK) but must at the same time acknowledge that some of these factors stem from the work and ambitions of the European Union since Sweden is a member of EU. Examples of this include the Payment Services Directives which aim at creating an internal, single market for payment services in the EU, which in turn has made European banks cooperate to the same effect in the form of SEPA (Single Euro Payments Area) and European Payments Council (EPC). EU is clearly of great importance for the Swedish payment system, but the analysis in this study still focuses on cash-based payments in SEK in Sweden. My study also focuses on developments in the time period 2012 – 2014 which is the time frame where cash payments started to decline in a more rapid pace than before. It is therefore a time period where many fundamental changes and debates related to a decline of the use of cash started. It should be noted, however, that the decline of the use of cash in Sweden became even more evident in the years 2015-2106.
References


Lagen 1988:1385 om Sveriges riksbank.


9: Dynamics: in the ICT Industry
Vicky Long

This chapter deals with a fundamental issue of Information and Communication Technologies (ICTs) dynamics: the industrial knowledge of ICTs. Four main knowledge approaches are introduced, in conjunction with a discussion on the extent to which these approaches contribute to our understanding of ICT innovation and diffusion. Three cases are further used to illustrate this linkage – between knowledge traits and innovation – based on evidence at both the sectoral and the micro (firm) levels.

Introduction

As one of the most dynamic industries today, Information and Communication Technologies (ICTs) has evolved rapidly, particularly over the last five decades, with an explosion after the 1990s. This industry therefore deserves a focused chapter, namely a detailed elaboration from the viewpoint of industry transformation and innovation.

It is important to first to point out that, as an independent industry sector, ICTs embrace many layers (e.g. “layer model”, 2002; 2007), running from basic connectivity, such as switches and mobile base stations, to highly interactive and (big data based) social media services, such as Facebook, Wechat and Uber.

This coverage, namely the scale and scope of ICTs, in turn, suggests variety of innovation/diffusion patterns involved in the development of ICT products, process, and services. It is therefore interesting, as an industry case, to understand the dynamics of its transformation.

Second, and perhaps most importantly today, is the fact that (sub) ICTs function as a generic (enabling) technology, which is instrumental to the transformation of other industry sectors. Typically, this relates to microprocessors, computers, Internet, Mobile technologies, and emerging areas like Big Data and Cloud Computing.

In scholarly research, there has been a recent association of this generic function of ICTs with the term General Purpose Technologies (GPTs). This term came into being two decades ago and primarily serves for the economic analysis of the linkages between technical change and growth (Bresnahan and Trajtenberg, 1995; Helpman, 1998; Bresnahan, 2010).
A typical GPT is: a) widely used, b) capable of on-going technical improvement, and c) enables innovation in application sectors. The combination of b and c has also been called “innovational complementarities” (Bresnahan, 2010, p764). The classic GPTs examples are: steam engines, electricity, combustion engines, computers, and Internet.

Looking further back, the generic function of ICTs has actually long been addressed in classic writings such as:

a) *Three Industry Revolutions* literature (Landes, 1969; Mokyr, 1990; Perez, 2002), which suggests that “the third industrial revolution, which is unfolding now, is fuelled by computers and networks.” (Brynjolfsson and McAfee, 2012, p.8);

b) *Information Age* literature (Castell, 1996; 1997; 1998), where it is argued that the re-configuration of a network society is enabled essentially by information flows;

c) *ICT Revolution* literature (Cohen et al., 2004), in which a digital divide (i.e. divergence in both ICT input and output) was brought to light;

d) *ICT Paradigm* literature, in which ICTs are associated with far-researching paradigmatic change in society (Freeman, 2007);

e) *Globalization* literature, ranging from *The Borderless World* (Ohmae, 1990) that highlights the dominance of global consumers, to *The World is Flat* (Friedman, 2005) that addresses how the companies, communities, governments, and individuals must adapt in a globalized society (to the ICTs era);

f) The (recent) *Second Machine Age* literature (Brynjolfsson and McAfee, 2014; Frey et al., 2013), in which the increasing digitalization and its impact on growth, on (un) employment, and on global convergence/divergence is brought to the fore.

To understand the very nature of the dynamics of ICTs, it is important to understand some fundamentals, among which is the industrial knowledge traits dimension. The forms of (sub-) ICT knowledge correlate to the particular patterns of change and transformation that have occurred in this industry sector.
The theme of this chapter is therefore formulated according to the following questions:

a) What are the categorizations of industry knowledge that are pertinent to ICTs?
b) How can these categorizations contribute to our understanding of ICT innovation and diffusion?

The chapter is structured as follows: section 2 summarizes the key empirical patterns of changes; section 3 introduces the knowledge traits approaches pertinent to ICT innovation and diffusion; section 4 illustrates, with three cases, the actual analysability of the knowledge traits approaches; and section 4 briefly discusses the managerial and policy implications.

**Setting the scene: ICTs in innovation and diffusion**

This section is an attempt to map out where we are now and what kinds of changes (possibly paradigmatic) have occurred in ICTs, both in terms of their development and their penetration into other industry sectors.

**Five decades of digitalization**

The civilization of mankind has experienced *five decades of digitalization*, starting approximately from the 1970s. If the first two decades (i.e. 1970s and 1980s) consist of technological breakthroughs in major sub-ICTs (e.g. Integrated Circuit), the last three decades (i.e. 1990s-) may be primarily characterized as digitalization, namely a wide usage (B2B) and consumption (B2C) of ICTs.

The 1970s was the era of microprocessors and IC (integrated circuits) development, which conditions high bit rate digital processing and transmission. A parallel development was the introduction of optical fibre technologies, thus enabling information to be carried through light.

The 1980s was the local area networks (LAN) era and inter-computer data transmission became more rapid. A parallel development was Cellular phone systems that were commercially tested and came into use (though not yet widespread).

The 1990s was the era of wireless LAN, personal area networks (PAN), and GSM (Global System for Mobile Communications) which started to become a *dominant design* (*Utterback* and *Abernathy*, 1975) in mobile communication. *Dominant design*, an evolutionary term, refers to key technological features that have managed to become a de facto standard (in contrast to a *de jure* standard).
The 2000s was the era of 3G mobile applications, which constituted quantitative breakthroughs in application and contents.

The 2010s is the era of the Internet of Things (e.g. smart grid application), going beyond the earlier machine-to-machine (M2M). It is also an era of Big Data, Cloud Computing, and a rapid development of Artificial Intelligence (robotization) (cf. Silver et al., 2016)

From a technology trajectory—e.g. paradigm shift—perspective, the above-mentioned ICT development milestones indicate waves of sub-ICTs development, on the one hand, and a convergence into all-in-IP ubiquity, on the other. Each wave consists of Schumpeterian bursts of innovation and strives for dominant design throughout the development process.

**Three directions of change**

Three directions of ICT development and diffusion may be identified:

The first direction is a deepening of digitalization throughout nearly all business sectors, and especially in traditional manufacturing (e.g. automobiles and other consumer electronics).

Design related industries are going through evolutionary, if not revolutionary changes. For example, visualization facilitates chassis simulation in Volvo cars; Additive manufacturing—also referred as industrial 3D printing—is used in Siemens gas turbines; Cybernetics proceeds in conjunction with, for example, Google's development in driverless cars.

This is a trend that, a decade ago, was often phrased as “industry/technologies convergence” and it primarily occurred in sub-ICT-sectors, e.g. between voice and data communication, or entertainment and media, content, and service provisions (e.g. payment apps) (Bohlin et al., 2000; Dixit, 2006).

Today the depth and breadth of the penetration of ICTs is different. Technologies—and their sub-modules—can be complemented and converged vertically across architectural layers, as well as horizontally across industrial value chains, as illustrated with the entries of Google and Apple into automobile industry. This development (theoretically) also opens numerous windows of opportunities in thus far unchartered fields, for example a robot surgeon or a robot scientist.

The second direction is a dramatic increase in user involvement in, for example, the development of applications, causing an exponential growth of information (perhaps un-sorted) in society.

This is exemplified by Facebook, YouTube, and Airbnb as participatory access platforms where consumers are also the main suppliers/contributors of
information and services. This is a consumer role that goes far beyond the notion of a passive receiver (e.g. cable-TV), as was the case in earlier decades. Seventy billion photos, as described by Brynjolfsson & McAfee (2014), are uploaded to Facebook each year (2013 data).

*The third direction is a fundamental change in the governance structure in network-based service provision.*

Since the 1980s, there has been a *global deregulation* in infrastructure-related service provision, for example in telecommunication service and in transport service provision. The depth and the breadth of that deregulation, however, vary across regions and industry sectors (cf. Horwitz, 1991).

Consequently, the “visible hand” of the state is mainly seen in regulatory tasks, for example in Telecom spectrum allocation (by National Regulatory Authorities), and in (industrial) policies such as innovation capacity programs (by innovation agencies).

Meanwhile, new and more diversified actors have started to have a say in the governance structure of networks, systems, devices, and applications. A multistakeholder co-regulating phenomenon is emerging, referred to by Brown and Marsden (2013) as an increasing “multistakeholderization [of Internet governance].”

Why has this multistakeholderization phenomenon occurred? I argue that three key technological developments may have contributed to the *multistakeholderization [of Internet governance] phenomenon:*

a) Development in, for example, dynamic spectrum access (DSA) technologies: this allows devices and/or applications to use a spectrum dynamically, namely when the spectrum is not used by another device or application, at a particular time, or in a particular geographic area\(^{62}\). That is to say, there is an increasing optimization at the technology layer that expands the capacity of utilization;

b) Internet protocols (IP) have become the common thread (for Internet governance). Skype and Wechat (for example, in WiFi form) go beyond mobile communication. Bitcoin is another example that may one day replace traditional banks as institutions that have played an important role in the history of industrialization of continental Europe, for example (Gerschenkron, 1963). While the role of the state should arguably be re-addressed in an era of a projected “green” transformation (Mazzucato, 2013), the state’s importance in Internet

\(^{62}\) Long and Laestadius (2016) discussed this feature in the Chinese TD-SCDMA mobile technology standard case.
governance is evidently fading. A co-regulation (by multistakeholders) governance structure is the future paradigm;

c) The projected boom of the Internet of Things (IoT): there are 1-2 billion machines/things to be connected by 2020 (ITU, 2015; 2016). A wide variety of devices and applications whose stakeholders have different capabilities, aims, and expectations (of others), are envisaged to be integrated into a uniform network of “things”. This calls for an increasing standardization to meet challenges in, for example, interoperability (technologically as well as institutionally). The evolutionary path is not going to be linear.

Regularities understood

There are at least two regularities (sometimes also referred as “laws”) that have been identified in the development of ICTs, which are closely related to each other.

Moore’s law

This law is originated from Gordon Moore (co-founder of Intel) who observed that every 12 months the number of transistors in a minimum-cost IC (Integrated Circuit) doubles (Moore, 1965). To simplify, there is a (memory/storage) capacity doubling every 12 months.

Variations of Moore’s law have been applied to improvements over time in disk drive capacity, display resolution, and bandwidth (Brynjolsson and McAfee, 2012, p3). This has long been considered the regularity in ICT development, enabling fast growth also in other industry sectors that use computing technologies.

“The second half of the chessboard”

This regularity mainly refers to an unnoticed exponential growth effect (Kurzweil, 1999; Brynjolsson and McAfee, 2012). It was originally referred to as the wheat and chessboard problem in the computer scientist Kurzweil’s (1999) elaboration: having one grain placed upon the first square of a chessboard and then doubling the number of grains on each subsequent square and continuing that way (i.e. constant doubling). What occurs in the second half of the chessboard (with this exponential growth) is a pile perhaps bigger than Mount Everest.

The point is that the consequence of technological progress initially goes unnoticed. The recent debate on our growing dependence on Artificial Intelligence (AI) typically illustrates this concern.


**ICT knowledge, innovation and diffusion**

Having gone through the stylized facts of the development and penetration (to other industry sectors) of ICTs in section 1, we now move a step further to understand why the above-mentioned scale and scope of changes have occurred. This is done by a knowledge traits elaboration, namely on the properties and characteristics of ICT knowledge. Four industrial knowledge approaches that are relevant to ICTs are presented and discussed in relation to ICT innovation and diffusion.

**ICTs embrace architectural as well as modular innovation**

Traditionally, innovation is categorized as either incremental or radical. To understand the key characteristics of the innovations that underlie the competitiveness of firms, Henderson and Clark (1990) suggest an additional break-down of elements involved in innovation, namely innovations that can be either architectural or modular.

*Modular innovation* refers to changes mostly in the components part (of a product). *Architectural innovation* refers to changes in “the ways that they are integrated into the system”, namely knowledge that is “embedded in the structure and information-processing procedures” (ibid., p9).

This dichotomy is supposed to help us to understand the “grand” and the “small”, the cumulative and the radical, and the *reinforced and overturned* innovations. As argued by Henderson and Clark (1990), established firms have difficulties in recognizing (and hence correcting) the errors in existing systems/architectures that are, in nature, embedded in their existing operations.

In addition to being useful for competitiveness analyses (of firms), this typology, as I see it, is also important to understand the varieties of innovations that have occurred in ICTs at both the system (architectural) level and the components (modular) level.

Two additional concepts may be brought up here to illustrate the extent to which the *architectural versus modular innovation* dichotomy can help us in understanding ICT innovations: a) Martin Fransman’s (2002; 2007) layer model; and b) the concept of modularity-in-design (Baldwin and Clark, 2000).

*Martin Fransman’s layer model*

Although ICTs are often looked upon and analysed as seamless technological systems (cf. Bohlin et al, 2000), it may be argued that they can be understood as a set of multifaceted physical layers, as illustrated in Table 1.
Tab. 1 The typology of ICTs. Source: Adapted from Fransman’s layer model (2002)

<table>
<thead>
<tr>
<th>Layers</th>
<th>Activities</th>
<th>Firms</th>
</tr>
</thead>
<tbody>
<tr>
<td>VI</td>
<td>Customer, Consuming</td>
<td>Spotify, YouTube, Airbnb, Uber</td>
</tr>
</tbody>
</table>
| V      | Application layer, incl. content packaging  
          (e.g. Web design, online information service, broadcasting service, e-commerce) | Facebook, Ebay, Amazon, Alibaba, MSN, Skype, Doodle |
| IV     | Navigation & Middleware layer  
          (e.g. browsers, portals, search engines, directory assistance, security, payment) | Yahoo, Google, Baidu |
| III    | Connectivity Layer  
          (e.g. Internet Access, Web hosting) | ISPs, IAPs |
|        | TCP/IP Interface |       |
| II     | Network layer  
          (e.g. optical fibre, mobile network, DSLlocal network, ISDN, ATM) | AT&T, Vodafone, Telenor, China Mobile |
| I      | Equipment & software layer  
          (e.g. switches, transmission equipment, billing software) | Ericsson, Huawei, Cisco |
This layer model aimed to provide a crude guide towards understanding the different kinds of industrial activities involved in ICT product creation and service provision (the boundaries between layers are not always clear, though).

Why is this layer model useful here?

One way to incorporate this layer model into the architectural versus modular innovation categorization is to view the general communication system(s) as a single abstract function. Each of the above mentioned layers then delivers a particular sub-function. In this way, each layer is a (giant) module. The selection of the modular, as well as the way(s) to organize them, is then an empirical question, which depends on a specific architecture to be delivered (to fulfil certain functions).

Another way to incorporate this layer model into the architectural versus modular innovation categorization is to view each layer as a separate system, an independent architecture with, for example, its own standard in interaction. It then gets its own (sub-) modular—invisible in the above table—to deliver that (layer) function.

Competition can occur within the layer as well as across layers. As I see it, what constitutes architectural vs. modular innovation is more of an empirical question rather than a theoretical abstraction.

While Henderson and Clark (1990) suggest that radical innovations often occur in modular, my view is “it depends”. A system (an architecture), having changed (out) more than 50% of the old modular and now delivering qualitatively different functions, would be classified as a radical innovation. My point is that there is no clear boundary and quite often, ICT innovations embrace elements of both architectural and modular innovations.

The next question is: how can we know which module should be changed and which one should be kept? This brings us to the second concept: modularity-in-design.

Modularity-in-design was originally introduced by Baldwin and Clark (2000) in their book Design Rules. This concept may also be viewed as an extension of Kim Clark’s earlier work on architectural versus modular innovation dichotomy (the same author).

Modularity-in-design

Modularity refers to an attribute of a product, or alternatively of a production process, that is capable of being “modularized”. To elaborate, the elements of a design, for example, can be split up and assigned to different module.
Modularity, according to Baldwin and Clark (2000) typically has three key advantages: a) to make complexity manageable, b) to enable parallel work, and c) to accommodate future uncertainty (ibid., p175).

For ICTs, modularity remains an important feature for chip design (Ernst, 2005), for open innovation (Chesbrough, 2003), and for business system integration (Davies, 2003). It is also important to explore its function in promoting new architecture creation in ICTs as well as in large technological systems (discussed by Petter and Pär in another chapter of this book).

To exemplify, Long and Laestadius (2016), through a detailed examination of a mobile communication system technology, found that modularity-in-design functioned as an important (technological) condition for late-coming developers in order to catch up. The developers (firms) could, in the new system creation, re-module “bits and bytes” in structured forms, as well as in fragmented packages, without having to grasp the wholeness of all the other systems (Long & Laestadius, 2016).

Empirically, this dichotomy—architectural versus modular knowledge—in ICTs can be further tested, not only in the ICT industry itself, but also in application areas (see section 3.1 and 3.2 cases).

**ICTs embrace analytical as well as synthetic knowledge**

This is a knowledge dichotomy in which it is important to distinguish the sectoral differences in industrial activities, on the one hand, and the geographical transferability of different kinds of industrial knowledge, on the other.

Following a tradition initiated by a Nobel Prize Laureate Herbert Simon (1962; 1969) on the distinction between the natural and the artificial in the artificial sciences, and by Faulkner (1994) on the differences between knowledge related to “experimental research” and knowledge related to “design practice”, Laestadius (1998; 2000) developed a twin concept: analytical-versus-synthetic knowledge. Industry sectors often display a different mix of these two kinds of industrial knowledge.

According to Laestadius et al. (2008, p256), “Analytical knowledge is based on deductive logic and is focused on deconstruction (“analysis”) of complex structures to increase the understanding of them”. This is a kind of knowledge that is “to a large extent codified” (ibid.), close to “know-why” (Moodyson et al., 2008) and dominating in, for example, the nanotech, biotech, and chemistry industries that essentially deduce science-related domain knowledge.
Synthetic knowledge, on the contrary, is “related to combinatorial skills” to integrate “technologies and systems into (more) complex systems or second-order technologies (designing artificial worlds)”, and essentially about achieving a “fit” between entities (Laestadius et al., 2008, p257), as in the polytechnics. Synthetic knowledge boils down to concrete technical problem solving (Moodyson et al., 2008). It is, comparably, more implicit (skills-based), process oriented, and inductive in nature (engineering based), and largely seen in process industries (in the pulp and paper industry, for example).

For the geographical analysis of industrial knowledge, this distinction has later been expanded into three types of knowledge bases, namely analytical (also called “science-based”), synthetic (also called “engineering-based”), and symbolic (also called “art-based”) (Asheim and Gertler, 2005; Asheim and Coenen, 2005; Asheim et al., 2007; Asheim et al., 2011; Martin 2012; Tether et al., 2012).

In geography studies, this categorization is mainly used to address the geographical “slippery versus stickiness” of the different kinds of industry knowledge. The actual balance between analytical and synthetic knowledge in different industries looks different. Generally, analytical knowledge is more geographically “slippery” than synthetic knowledge.

Why is this knowledge taxonomy relevant to the innovation and diffusion of ICTs?

Long (2014), in her doctoral thesis, brings this knowledge taxonomy into the understanding of ICTs, in relation to the learning patterns across regions.

Long’s (2014) research started from a crude categorization of “ICT manufacturing” as synthetic knowledge and “ICT software” as analytical knowledge. With the support of statistical data, Long (2014) confirmed the relative geographical slipperiness of analytical knowledge over synthetic knowledge within ICTs (in transferring), on the one hand, and argued that the existence of the right sort of absorptive capacity at the receipt side is an additional necessity.

Continuing with Long’s (2014) line of research, we may move on to the statistical definition of ICTs:

A statistical breakdown of the International Standard Industrial Classification’s (ISIC) definition on ICTs (Rev. 4) is as follows, with industry codes (“hierarchical, four-level structure of mutually exclusive categories”):

- ICT manufacturing industries (2610, 2620, 2630, 2640, 2680)
- ICT trade industries (4651, 4652)
• ICT services industries
• Software publishing (5820: 5821, 5829)
• Telecommunication (6110, 6120, 6130, 6190)
• Information technology service activities (6201, 6202, 6209)
• Web portals, data processing, hosting and related activities (6311, 6312)
• Repair of computers and communication equipment (9511, 9512)

Evidently this is an (engineering) activities-based definition. It is therefore not far-fetched to connect this definition with the industrial knowledge taxonomy mentioned above. Difficulties remain, for example, concerning where we can position “ICT trade industries” in the knowledge taxonomy. Nevertheless, as Long (2014) exemplified, this taxonomy is important to understand the geographical patterns in, for example, ICT technological knowledge diffusion.

2.3 ICT knowledge is complex rather than discrete

This categorization—complex versus discrete technologies (Nelson and Merge 1990)—may also be considered as an alternative to the traditional view of categorizing innovation as either incremental or radical.

In 1990, Richard Nelson (an evolutionary economist) and his colleague Robert Merge (a law professor), by looking at the scope of patents, established a distinction between two generic types of technological industries, namely complex and discrete technologies.

ICTs typically belong to the “complex technologies” category, in which an invention often has many potentially patentable elements and they are interlinked. As a consequence, an ICT patent is often “standing on the shoulders of giants”, as Isaac Newton puts it, which Suzanne Scotchmer (1991) also referred to in her analysis of the scope of patents and patent citations.

On the contrary, within “discrete technologies” that are typified by chemistry, for example, the inter-connection within an invention is not as strong as in the opposite category. To elaborate, the number of potentially patentable elements within an invention is limited. As a consequence, one chemical patent, for example, is often a separate invention from another.

This distinction—complex (cumulative) versus discrete technologies—is widely adopted and used in analysing appropriability conditions (Cohen et al., 2000; Kash and Kingston, 2001), and in analysing corporate IP (Intellectual Patents) strategies (Reitzig, 2004). Appropriability refers to (environmental) factors that govern an innovator’s capability to capture returns from innovations (Teece, 1986).
Beyond that, how can this dichotomy—complex versus discrete technologies—contribute to our understanding of the patterns of innovation and diffusion in ICTs? Two examples may be brought up as follows.

Firstly, it helps to understand innovation. The Schumpeterian “bursts of innovations” within ICTs are “complex”. As Nelson and Merge (1990) put it, a new ICT innovation is often “standing on the shoulders” of a prior ICT innovation. This implies the ICTs often have a generational character. That is to say, there are generations of new technologies and each step (generation) signals a qualitative breakthrough, as well as a quantitative accumulation in knowledge.

For example, mobile system technologies are often labelled as “2G”, “3G”, “4G”, and “5G” (visible also in mobile handsets in Sweden). “G” refers to generation. In the case of mobile system technologies, the evolution—or one path of the evolution—has gone through analogue communication technologies (1G) to digital technologies GSM (2G), and then to WCDMA (3G) and to LTE (4G), and so on (cf. Long and Lasteadius, 2016).

Does this imply that an ICT (cumulative) innovation is infinite, possibly with a long tail? The answer may be “not necessarily”.

There are deadlocked (or dead-end) ICT technologies, such as the once widespread Personal Handy-phone System (PHS). Originally developed by NTT DoCoMo in Japan at the end of 1980s, this technology managed to tag along more than 100 million subscribers in Asian countries during its peak (around 2004) (Long & Laestadius, 2016).

The cumulative nature of ICT innovation, however, may imply a prolongation of the product life cycle (of an old ICT product/technology) in practice. For example, in Ericsson, GSM (Global System for Mobile Communication, a 2G technology), within the first decade of the 21 century, still gained many cumulative innovations through collaborations in 3GPP2 (The 3rd Generation Partnership Project 2). This creates more challenges for any later innovations. That is to say, a later innovation has to have substantial advantages over the prior innovation (that also undergoes continuous improvement) in order to take over the market.

The cumulative nature of ICT innovation also has great implications for innovative new entries, e.g. start-up firms. The incumbent actors, often patentees in a prior innovation, “could eliminate competition in the form of substitute technologies through fencing” (Reitzig, 2004, p457) (in, for example, patents). That, in turn, constitutes a kind of “patent thicket” (Shapiro, 2001) in which a late-coming entry has difficulties going through.
Secondly it helps understand knowledge appropriation. The word “appropriation” often connotes “profiting from innovation” in Teece (1986) terms.

For example, we see many forms of collaboration among ICT firms in technology standards (development consortia). In practice, one ICT invention is hardly standing alone due to the cumulative nature of ICTs in innovation (as “complex technology”). Consequently, the forms of cooperation and competition are adapted. This is illustrated in “patent pooling” and alike industry activities in standardization. Problems such as “patent thickets” (Shapiro, 2003) that arise due to mutual blockages in patenting also occur. This is a kind of dilemma faced by managers and other practitioners within the industry.

Moreover, ICT embraces many sub-ICT-sectors in which the nature of the knowledge also differs. A patent is, for example, connected with Telecom capital goods supply and copyright is connected with software product/service. IC chips use patents and faddish consumer goods such as mobile terminals use market lead to appropriate returns. There are mixes of the different appropriability mechanisms across sub-ICT-sectors.

Moreover, the dynamics of user-generated contents (e.g. Facebook, YouTube), increasingly seen in ICT sectors, poses new challenges, particularly for firms to appropriate return and for institutions to re-set the incentive structure to encourage entrepreneurship and innovation.

**ICT knowledge is both infrastructural and proprietary**

This categorization also touches mostly on the issue of knowledge appropriation, namely how firms profit from (ICT) innovations.

The general problem formulation, i.e. that knowledge is both infrastructural and proprietary, may be traced back to the so-called “information paradox” that was raised by Nobile Laureate Kenneth Arrow (1962) in his innovation incentive discussion. To put it simply, a large proportion of the information/knowledge has to be transferred to customers to ensure that values (of the information/knowledge) are being perceived and accepted prior to the purchase. However, once this (knowledge transferring) process is completed, there is no incentive for the customers to pay for that knowledge any longer as it has already been acquired. Arrow’s conceptualization of the information paradox is largely used to argue the need for patent protection in innovation.

Inspired by Arrow, Steinmueller (1995; 2002), in his competition analysis, argued that ICT knowledge, in particular, is characterized as both infrastructural and proprietary.
“ICT knowledge being infrastructural” connotes: a) the General Purpose Technologies (GPTs) dimension of (sub) ICTs, b) the cumulativeness of ICT knowledge discussed in section 2.3, and c) the need for collaboration in ICT innovation.

“ICT knowledge being proprietary” means that viable business models need to be created to keep incentives in place. Innovation incentives are discussed at two levels. At the society level, both patent and copyright and alike institutions are created for sustaining incentives for innovation. At the firm level, it is essentially an issue of strategy for how to profit from innovations, “appropriability strategy” as Teece (1986) puts it. At the micro (firm) level, a patent is not necessarily considered the most efficient means to appropriate returns (Cohen et al., 2000), and the strategies vary a lot between industry sectors. For example, the pulp and paper process industry uses mostly trade secrecy, rather than patents, to appropriate returns.

At the first extreme, there are open innovation movements that are illustrated with open sources (Chesbrough, 2003; Pisano, 2006). Terms like “copyleft” illustrate this kind of ideological development. “Copyleft” is a pun on the word “copyright” and it advocates the obligation to preserve the same right on the derivative works. The right often refers to using and modifying others’ works free of charge.

At the second extreme, cases like “patent mining”, “patent blanketing”, “patent blocking”, namely over-patenting, illustrate the so-called “tragedy of the anti-commons” (Heller and Eisenberg, 1998). Where is the balance of patenting and blocking, for example? This issue continues to be of great interest, especially among industry managers and policy makers.

The notion of knowledge as infrastructural versus proprietary is taken and formulated in alternative ways, for example, in literature on sectoral systems of innovation (Malerba and Orsenigo 1996; Malerba, 2004), according to which technological opportunities and appropriability conditions differ greatly across industry sectors.

This categorization is important to understand why different means of knowledge appropriation exist across industry sectors.

**Case illustrations**

In the following, three cases are used to illustrate the analysability of the knowledge traits approaches. The units of analysis are different. In the first case, the focus is on a firm (firms); the second one focuses on a technology, and the third one on a service (rather than a product).
Nokia versus Huawei

This case serves to illustrate: a) the use of the *architectural versus modular knowledge* approach in viewing ICT innovation, and b) the linkages between a firm’s knowledge and its selection of the technological trajectory. The unit of analysis is a firm(s).

Nokia, compared to Huawei, is a forerunner in Telecom equipment supply and has dominated mobile terminal supply since the introduction of GSM (Global System for Mobile Communications) in the 1990s. GSM is a 2G (generation) technology—a “dominant design” according to Utterback and Abernathy (1975)—and was considered a radical step that evolved from the analogue based 1G-communication technology. It is not exaggerating to state that, from the 1990s to the first half of the 2000s, Ericsson dominated the system technologies supply in GSM, while Nokia dominated in the second half.

It may be understood that the *architecture*—the GSM—remains dominant and unchanged, while both Ericsson and Nokia as industry actors were continuing to add new features, new *modules*, into the system. This is often done in collaboration within 3GPP (The *3rd Generation Partnership Project*) and alike industry consortia.

2G technologies gradually faded while 3G technologies—with higher bandwidth in data transfer and better facilitation with mobile applications—gradually took hold of the market. 2007 is the year that smart phones were introduced. It was also the year that Nokia started its downturn as a terminal provider. New actors like Apple—from the data-communication rather than mobile communication side—started to squeeze into the mobile market and gradually blossom.

Huawei, established in 1988, is a latecomer in Telecommunication supply (compared to Nokia and Ericsson), and a follower since GSM was introduced. What is different with Huawei is its projection of an “all-in-IP” trajectory in communication already at the end of 1990s. That is to say, data communication was given equal weight as telecommunication in a new *architectural* design at the early stage. Huawei’s strategy is elaborated in Long’s (2014) doctoral thesis, and also illustrated with Huawei’s partnership with 3com in 2003 in the field of Enterprise data Networking (to compete with Cisco). “All-in-IP” became the trajectory for mobile communication and smart phones gradually began to dominate the mobile terminal market. In the second half of the first decade into 21st century, the “new” entrants—Apple as well as Huawei—won the battle of mobile handsets.

While innovations can occur both at the architectural and modular levels, what is decisive to win the market is a question that is open for empirical study. While innovations are also largely path dependent, i.e. dependent on
developers’ prior knowledge, Huawei’s case illustrates that learning with an open approach could break out from a knowledge lock-in. That is to say, there are values of learning from actors positioned in different trajectories.

3.2 Smart grid and green ICT

This case serves to illustrate: a) the General Purpose Technologies (GPTs) function of ICTs, and b) the usage of the architectural versus modular knowledge approach in ICTs. The unit of analysis is a technology.

The Background

“Green ICTs” (European Commission, 2010) is an emerging area of debate (and research) that is primarily linked to the challenges triggered by climate change. It mainly addresses the low-carbon “green” effect of ICT technologies, sometimes also called “green informatics” (European Commission, 2010). ICTs, for example, in the EU 2020 project, are expected to contribute not only to innovation and growth, but also to the achievements of a low-carbon society, through raising energy efficiencies throughout the economic sectors, for instance.

Smart Grid—digital technologies (the “digital layer”) marrying with power grids—is a typical illustration of the “green” agenda and of the General Purpose Technologies (GPTs) function of ICT.

Questions:

How smart can a Smart Grid become with the help of ICTs?

To which extent can the architectural versus modular knowledge dichotomy (see section 2.1) help our understanding of this Smart Grid case?

Technologically, the facilitation of ICTs here can be seen on least at two fronts.

Firstly, and on the operational front, with the help of digital processing and digital monitoring (e.g. smart meters) technologies, the grid system can be more efficient in monitoring the (voltage) fluctuations of the electricity supply, whereas unstable renewable sources (e.g. wind/solar power based) take increasing proportions.

Using the jargon of architectural versus modular knowledge, we may say that ICT modular are added into the existing grid (architectural) systems. Old modular are replaced with new ones, with the overturned (Henderson and Clark, 1990) impact on the functionalities of the grid system.

Secondly, and on the information front, the real-time information (and communication) technologies re-define what was mentioned earlier about
reliability and security. As Einarsson and Svensson’s (2014) report pointed out, one typical planned addition to the existing Network Information System (NIS) (documenting, for example, cables and transformers) is the Geographic Information System (GIS) function. This, in turn, helps to identify also the “where” answer in a complex (grid) system.

Again, new modules (Henderson and Clark, 1990) such as GIS are integrated. The concept of modularity, as discussed in section 2.1, is highly relevant here.

Remaining on the theoretical level (without actual cases), Smart Grid is an interesting case to follow, particularly in terms of how far ICTs can facilitate the improvement of energy efficiency and, consequently, also contribute to a low carbon society. While the supply versus demand mix of the electric grid system looks different across countries, Smart Grid implementation is an issue beyond technology (e.g. ICTs) availability. It is open for empirical examination, for example concerning concrete problems of the deployment and diffusion of ICTs.

3.3 M-pesa

This is a case illustrating the General Purpose Technologies (GPTs) function of ICTs, on the one hand, and a service (rather than a product) innovation, on the other hand. The GPTs function of ICT may be perceived as an alternative formulation of ICTs being infrastructural (versus being proprietary) (see section 2.4). The unit of analysis is a (payment) service, or a platform, an institution, a “social technology” (rather than a “physical technology”) in Nelson’s (2008) terms.

Background A: the digital divide

Although accessing the digital world is no longer a privilege of a few, disparities remain between those who have access and those who do not. By the end of 2014, there were still 4.3 billion people who were not online (ITU, 2015), a fact easily forgotten by people living in OECD countries, for example. According to Hilbert (2016, p567), “the bad news is that the digital access divide is here to stay” if the digital access divide is measured by bandwidth rather than by number of subscriptions, which shows that “the gaps are rapidly closing” (cf. Compaine, 2001) over time.

Background B: the financial services divide

Financial services are taken for granted in the developed world. While the financial crisis of 2007–09 is still often discussed today in the context of an aimed recovery, for example in OECD countries, we often forget that it (the crisis) is also a “privilege” enjoyed by people in developed world. Many people in the poorest parts of the world simply do not have a bank account.
example, before the introduction of M-pesa in 2007, less than 20% of adult Kenyans had bank account (The Guardian, 2007).

M-pesa is a mobile phone-based platform for money transfer. It was launched in 2007, by a Mobile network service provider Safaricom (Vodafone’s African branch) in Kenya. It initially aimed to facilitate poor people who do not have a bank account, but do have a basic mobile phone, when transferring money to each other. Now it is widespread in many (African) countries. The underlying technological system was not sophisticated. It did (does) not require a smart phone as, needless to say, the bandwidth is often not sufficient in developing countries.

Today, M-pesa has totally transformed economic interaction in Kenya. In 2013, 43% of Kenya’s GDP flowed through M-pesa, with over 237 million transactions. Arguably, it is particularly well suited to low-income groups: in 2011, 72% of Kenyans living on less than $1.25 a day were using M-Pesa, an increase of 20% over 2008 (The Economist, 2015 Aug,12). As a business model, it later was expanded also into Tanzania, South Africa, and India.

M-pesa is therefore an interesting service case illustrating that ICT knowledge is both infrastructural and proprietary.

M-pesa marries ICTs into traditional banking services. This money-transfer platform is *infrastructural*, a character that many bank services also have. It is also *proprietary* in the sense that Safaricom is not giving out service free of charge. On the contrary, because of M-pesa’s very dominance in Kenya, Safaricom now charges fees for money transfers that are very high, i.e. up to 10% of the face value (The Economist, 2015 July). This reflects a connection to the *appropriability regime* concept discussed in section 2.4: the extent to which firms can make money from the product/service offerings hinges on the character and knowledge traits of the innovation/service, as well as the landscape of competition.

**Managerial and policy implications**

This chapter is based on evidence at the sectoral and at the micro (firm) levels. Many of the managerial implications have already been discussed in earlier texts. To summarize, the knowledge traits approach helps us to understand: a) the very causes of (ICT) firms’ competitiveness, and b) why phenomena, such as collaborative competition in, for example, technology standards (patent pooling), are common in this industry.

On the policy implication front, in sections 2.3 and 2.4 the knowledge appropriation parts are the most relevant. Innovation policies/institutions
need to be finely tuned. For example, Suzanne Scotchmer (1991) suggests that the breadth of patent protection of a prior patent can be a determining factor for the existence of the second patent. This is particularly true in the category of complex technologies (section 2.3).

Is there an optimal incentive scheme for innovation or not? On the one hand, the existence of IPR institutions, for example, is important to encourage innovation. On the other hand, this very protection is also part of the causes of “patent mining” and “patent blocking”, which are both extreme actions. Where is the balance?

Suggested further readings

Empirically, reports and studies related to digitalization and its various subsets (e.g. cloud computing) and its various applications (e.g. smart cities; smart grids) are all relevant. These are essential empirical grounds. Many of these writings, however, may still remain as industry reports at this stage, and therefore need to be critically assessed.

Theoretically, perspectives related to technological regime; (ICT) sectoral innovation systems and/or modules, such as open innovation; (ICT) regional clusters and knowledge spill-overs; (ICT) trajectories/paradigms; (ICT) diffusions and technology transfer, digital divide, and digital business are all relevant to certain extent. They shed light on our understanding of the different mechanisms involved in the dynamics of ICTs.

In a concrete analysis such as a master’s thesis, the adoption of the perspective(s) is a matter of choice that best highlights the empirical actualities.

On methodological considerations and the use of sources in this chapter

This is a chapter with empirical pattern extractions. It is also a chapter with analytical abridgements of earlier disconnected theoretical fields, such as ICT studies, innovation literature, and industrial knowledge base literature.

This is a chapter that embraces secondary sources (e.g. literature), as well as first hand data, namely the author’s own observations and research (and of course also reflections and interpretations).
References


Brynjolfsson, E., and McAfee, A. (2012). Race against the machine: how the digital revolution is accelerating innovation, driving productivity, and irreversibly transforming employment and the economy. The MIT Sloan School of Management.


Frey, C, B; Osborne, Michael (2013) The future of employment: how susceptible are jobs to computerization? Working paper by Oxford University Engineering Sciences Department. http://arche.depotoi.re/autoblogs/wwwinternetactunet_8a3fe3331e0ad7327e18d9fe6ec3f0ad04dcea58/media/3722fa7d.The_Future_of_Employment.pdf


Malerba, F; Orsenigo, L (1996) Schumpeterian patterns of innovation are technology-specific, Research Policy 25, 451-478


The Economist (2015, July 11th, ) A new east Africa campaign: Two African business giants go head-to-head over mobile telecoms and payments

The Economist, 2015 Aug,12) M-Pesa And The Rise Of The Global Mobile Money Market

The Guardian, 2007-03-31, Kenya sets world first with money transfers by mobile


Utterback, J; Abernathy, W (1975), A dynamic model of process and product innovation, Omega, 3 (6) 639-656
10: Structural Change in the Petroleum Industry

Maria Morgunova & Vladimir Kutcherov

In this chapter, we aim to contribute to the understanding of industrial dynamics through the concept of development blocks introduced by Dahmén (1950/1970) applied to the case of the petroleum industry (oil and natural gas). We first aim to offer some insight into the current state of the energy sector and the petroleum industry by describing the development of the energy system from the 1800s to the present day to reveal the system’s structural changes. Second, we focus on the definition of the development block. We suggest a brief overview of relevant notions and the manner in which they function within the conceptual framework. Third, we apply the development block concept to the case of the petroleum industry. We want to explore the new structural tensions and transformation pressures in the petroleum industry that have occurred within the last transformation period. Finally, we elucidate the theoretical and empirical conclusions both in terms of the use of the concept of development blocks and of the dynamics of the petroleum industry by focusing on the latest industrial period; then, we suggest managerial and policy implications deriving from our investigation.

The changing energy consumption paradigm

The energy sector includes all the industries involved in the extraction, production, refining, transportation and distribution of energy. The energy infrastructure and energy institutions are deeply embedded in the economy. Investments in the global energy supply are estimated at 1600 billion U.S. dollars annually over the 2011-2013 period; these expenditures have more than doubled since 2000 in real terms (IEA, 2014:20). Moreover, these investments include all energy activities and sources – from hydrocarbon fuels, power stations, and renewable energy sources to the means of processing and transporting energy sources – and reflect the growing energy demand and rising production costs. The petroleum industry is the second most capitalized sector in the global economy, following only the banking sector (Financial Times, 2015). Meanwhile, crude oil and natural gas cover 56.8% of total energy consumption (BP, 2016b). The energy sector is no longer
merely a business that supplies an important commodity; instead, its roots are deeply embedded in society.

The petroleum industry is a business with a much longer history than is often presumed. Its current status as a component of a complex socio-economic and political system comes from the early 20th century with the invention of the internal combustion engine. The petroleum industry passed through different development stages and began to influence business and politics, eventually entangling the global economy in the ups and downs of the crude oil price. At present, the industry may be entering a new stage in which dramatic changes in economics, geopolitics, technological progress and ecological conditions have combined to form a new business reality, which means “…business as usual is no longer a credible future for the oil and gas industry…” (Mitchell et al., 2012). With the aim of studying the underlying transformation in the petroleum industry, use the development blocks concept that is developed later in this chapter and to focus on the tensions and pressures that are responsible for this transformation.

In general, the constant transformation of the energy system yields better ways to utilize energy and introduces new types of energy sources. A historical analysis of energy consumption can help identify a number of changing and also challenging trends. Figure 1 represents global energy consumption by sources of energy (as a percentage) over the 1800-2035 period. It also shows the energy source that was dominant during a certain historical period. Biomass (or wood) was the major energy source for centuries, until the 1850s, which coal began to occupy a higher share of global energy consumption because of the spiraling “age of synergy” (Smil, 2005) or the second industrial (technological) revolution.
During this time, the energy intensity of national and international economies, production sectors and of society in general increased dramatically. Coal dominated the world energy market until the 1960s. At the beginning of the 1900s, crude oil first entered this market in the form of kerosene as a new lighting source. It took almost 60 years, until the 1960s, for crude oil to occupy the biggest share of the world’s energy consumption (Podobnik, 2006:5). Increasing automobile production in the 1920s and global geopolitical conflicts such as World War I and World War II necessitated the shift to the ‘gusher age’ of more energy effective crude oil use. Natural gas’s share has also grown steadily in fuelling global energy demand together with crude oil; notably, in the first decade of the 21st century, its share of global energy consumption has almost drawn even with that of coal for the first time in history. Renewable energy sources, as a new sustainable way to provide energy to the world, occupied less than 1% of the energy market in 2000. According to forecasts (BP, 2015a, b) – the share of renewables will exceed 6% by 2035 (Figure 2).

In the most recent 20-year period, total energy consumption has risen almost by 40%. However, the shares of crude oil and natural gas in the energy balance are still roughly the same, and approximately 60% of global energy consumption is supplied by oil and gas (Figure 2) (BP, 2015c, 2016b).
Increasing energy consumption, growth in the energy industry, the impact and importance of the energy industry, and advancing technological development have accelerated transformations in the paradigm of energy consumption. Real volumes of energy consumption have also risen dramatically and have been successfully incorporated into the world economy. In each period of the ‘changing energy consumption paradigm’, the economy, which was initially based on one dominant source, was able to find a trade-off toward another energy source when consumption increased substantially, which meant “relative rather than absolute shift[s]” (Podobnik, 2006:7). However, crude oil, natural gas and coal are still projected to supply the highest share of energy consumption into the foreseeable future, and crude oil is the unquestionable leader through the present. Other energy types are far from such rapid growth trends (see figure 1). In this vein, the transformation of the petroleum industry, meaning the rapid expansion of crude oil and natural gas production (and consumption), is under scrutiny. We suppose that the origin of that transformation is very much dependent on the origin of the structural tensions and transformation pressures in the petroleum industry – concepts described by Dahmén (1950/1970). Thus, we intend to apply transformation theory to study the petroleum industry, but we must first understand the main foundational aspects of the Dahménian development blocks concept.
The development blocks concept and what it explores

Industrial dynamics originates from systems approaches, has roots in Marshall’s theories, and has been developed in the works of Schumpeter. Interest in industrial dynamics was developed by Dahmén (1950/1970), Forrester (1961 and 1968), Klein (1977 and 1984), (Carlsson, 2008), and came into the modern era with Carlsson (1987, 1989) and others (e.g., Geels, 2007; Geels, Kemp, 2007). The earliest efforts involved in adding to established theoretical constructs in the economics literature was undertaken by Forrester (1968) in describing the state and development of the field after the first decade it existed, and Dahmén (1950/1970, 1988) in defining development blocks in industrial economics and incorporating historical analyses into it. The development of a new approach to system studies was supported by the evolution of modern knowledge-intensive industries, such as biotechnology and telecommunications (Krafft, 2006:17). Industrial dynamics and areas of research have been explicitly discussed in previous chapters, and a detailed discussion of Dahménian approaches is undertaken by Laestadius’ chapter in this volume. In this chapter, we will use Dahmén’s theory on development blocks and apply it to the empirical case of the petroleum industry.

Dahmén (1988:3-4) was primarily interested in the transformation of industry and trade and not just in resources allocation or the subjects explored by macroeconomics (e.g., business cycles, employment and price levels). Dahmén’s doctoral thesis, as described by Laestadius' chapter in this volume, targeted entrepreneurial activity in Swedish industry (Dahmén, 1950/1970). Dahmén investigates industrial transformation in a stepwise manner, which roughly can be divided into an introduction of something new and a scrapping of the ‘old’ parts of a system. On one hand, there are new methods of production and marketing, new products and services, new markets themselves, new resources and energy, and, on the other hand, there are bygone methods, products, services, markets and resources. Thus, a key concept in all the Dahménian studies is transformation. According to Carlsson (1989:3), this was probably the first study that addressed Schumpeterian analysis, and it focused on entrepreneurship and the tension between existing resources and capabilities and on using them for the sake of new business opportunities.

In his doctoral thesis, Dahmén introduced the concept of development blocks that we are focusing on in this chapter. Development blocks, as a new framework and basis for industrial transformation studies, were defined as “sequences of complementarities which by way of a series of structural tensions, i.e., disequilibria, may result in a balanced situation” (Dahmén 1989, p. III). These development blocks underline the “co-evolution of parts of the economy” (Enflo et al., 2008) in which transformational processes are driven by innovations and new technologies. Complementarities of a different
kind, or co-evolution parts, include institutions, companies, and entrepreneurs within a development block, where they have production-related connections. The concept is illustrated in Figure 3 and describes the transformation process of a development block from a disequilibrium (I) to a relatively balanced position (II) and again to the next stage of disequilibrium (III). Figure 3 also provides an overview of the development block components, inner and outer pressures and tensions, which we discuss in detail below. We seek to underline that Dahmén’s development blocks concept embodies a systems approach and must be treated in that manner. With this introduction, let us begin the discussion of how development blocks are is formed and transformed.

![Figure 3: The transformation process of a development block.](image)

As a system, it is assumed that a development block always strives towards a balance. However, the development block is in fact never in a complete balance, but is rather in constant transformation with a few periods of relative balance. There are several reasons that the system will fall into disequilibrium. The development block itself comprises different types of complementarities, such as industries, companies, institutions and other organizations. First, when the complementarities do not match the production or development pace of the other complementarities, there is disequilibrium in the development block. Another reason for disequilibrium involves outer transformation pressures that force the system to transform and adapt. The transformation pressures can be initiated by the following phenomena (Carlsson, Henriksson, 1991:127-128, 137): “(1) introduction of new methods of producing and of marketing products and services; (2) appearance of new and marketable products and services; (3) opening up of new markets; (4) exploitation of new sources of raw materials and energy; (5) scrapping of “old” methods of producing and marketing products and services; (6) disappearance of “old” products and services; (7) decline of “old” markets; (8) closing of “old” sources of raw material and energy; and also (9) formation of new “institutions”, i.e., political and organizational structures and systems, legal framework and compliance procedures, etc.” Finally, the
lack of balance is itself a driver of transformation towards a more balanced position, which leads to structural tensions within the development block. Thus, the transformation process sometimes requires a coherent development of complementarities (Dahmén, 1989). The structural tension can have a positive and a negative character with regard to the system’s development (i.e., technological updates and innovativeness), and is described as existing on three levels. The first level appears on the technological level, such as bottlenecks. The second level is on the industrial level, where markets/technologies/methods of production do not fit together. The third and final tension level appears on the societal level, where industrial sectors may not well fit together.

Simply put, the transformations can be summarized into an introduction of something new and a scrapping of the ‘old’ parts of a system, as previously mentioned. As structural tensions can have either a positive and negative character, transformation can also be positive or negative. Positive transformation creates new opportunities by inventing something new. Conversely, negative transformation within the development block has its roots in Schumpeterian concepts, such as “creative destruction” (Schumpeter, 1942), as it brings the “necessity to adjust and to adapt” (Carlsson, Henriksson, 1991). The character of the transformation process may depend on the type of structural tension that brings disequilibrium. The development block can also change its character during its transformation, such that system behaviour will also be changed. The historical examples of some of those transformations initiated partly by social and political (and also economical) tensions include the US depression during the 1930s, the immediate aftermath of the Second World War, etc. Some industrial and technological crises in which development blocks theory was successfully applied include the ICT boom and crisis in the 1990s and 2000s (particularly as it occurred in the Nordic countries, i.e., Sweden and Finland) (Erixon, 2011:124). In addition, structural tensions are repetitive phenomena (Carlsson, Henriksson, 1991). In principle, the character of those tensions is similar; however, they appear in different settings. In periods of drastic transformations, such as revolutions and other critical events (such as wars), the imbalances grow vigorously and bottlenecks appear. However, a platform for implementation of changes is also being created.

The gap fillers on the way to a relatively balanced system include innovations, which has been illustrated in Figure 3 – the relative balance (II) stage. The outer and inner tension gaps in the development block are to be filled by innovations or a new technology, which might stabilize the development block. Those can appear both within and outside of the system. The development block concept is in general very much based on technological innovations and entrepreneurship. Innovations create new links within the development block – including new processes, functions, and methods of
production among complementarities (Enflo et al., 2008). This process by itself, whether it is positive or negative, leads to a constant imbalance within the development block. For example, innovations motivate entrepreneurs to search for new entrepreneurial activities, which can lead both to positive and negative transformations. Along with new transformation pressures and appearing structural tensions, these features can bring the system further into the next disequilibrium in a transformation process. However, the impact of new technologies is almost impossible to access in full because of system complexity. It also must be mentioned, that development blocks are not synonyms with innovation clusters (Erixon, 2011). Connections in development blocks are not only due to innovations but also to investments and production-related processes.

There are some additional characteristics of the development blocks and related notions. First, within a development block, the more progressive industries are those that are on the positive side of the transformation process, which means using innovative technologies to increase production share, opening new opportunities for industrial activities, etc. (Erixon, 2011; Carlsson, Henriksson, 1991). Thus, in general, the progressive development of a block would involve its extension that was based on progressive industries and their complementary input-output relations with other industries (Dahmén, 1989, 1991a,b). Continuing the discussion, another side of this characteristic is that extension can cause the structural tensions in the development block (or “creative destruction”), as other industries may be lagging behind. Second, development blocks are much more effectively formed in decentralized – as opposed to centralized – structures, and this process can also be accelerated by market pressures (Carlsson, Henriksson, 1991). Third, development blocks can also be classified as intentional and unintentional, organized by entrepreneurial groups or independently (Carlsson, Henriksson, 1991). These three qualities impact the actual analysis of motives, intentions and tensions that arise.

In general, the development blocks theory is a way to incorporate macroeconomics and the institutional dimension into business cycle theories (Erixon, 2011). In other words, it helps explore the effects of technological and institutional changes separately from economic development, and it makes the analysis methodologically and empirically stronger. Other authors describe the task of the development block as merging the analysis of economic history, geography and theory (Carlsson, Henriksson, 1991). Dahmén and his mentor Åkerman included endogenous changes, societal factors, the composition of industries, relations among industries and companies, and also a within-country perspective. The concept of development blocks also acts as a bridge to the analysis of structures, cycles and growth (Erixon, 2011:120). Erixon (2011) describes their approach as qualitative and case-oriented, because it was mainly based on defined periods
of industrial development in certain countries to identify driving forces and structural changes (Åkerman, 1946, 1951, 1960; Dahmén, 1950/1970). Among the key driving forces and factors of such industrial transformations are political and institutional forces and factors. However, the pivotal role for Dahmén and Åkerman was always played by technological progress based on innovations. Critics of Åkerman-Dahmén theoretical development (with regard to both development blocks and, of their general business cycle theory) Erixon, 2011:121-122) mainly reference a number of issues, such as the diffusion of innovations over the recovery periods of industries, the little attention paid to macroeconomic conditions and policies, and the orientation towards large industrialized countries. Another important critic targets the contradiction between the “emphasizing role of progressive industries” (Erixon, 2011:123) and the fact that Sweden (the main country under investigation in Åkerman-Dahmén’s studies) in some historical periods benefited from traditional resource-based industries.

To conclude, the theory of development blocks is a good source of material for empirical investigations and a platform from which to engage in economic-historical research; on the other hand, it can also contribute to the theoretical discourse (Carlsson, Henriksson, 1991). In industrial economics, the development block concept has been applied to a number of interconnected factors in the development process (Carlsson, Henriksson, 1991), such as cost and price signals, new technology and entrepreneurship, etc. Both bottlenecks and entrepreneurial opportunities are being investigated to understand the character of industrial transformation processes and their influence on the corresponding industries and society. The focus on the character of transformations can also serve as analytical steps in the relevant studies (Carlsson, Henriksson, 1991:137), where the higher focus on structural tensions and their levels (technological, industrial or societal) will provide different cut-offs for the analysis.

The periods of transformation of the petroleum industry

The petroleum industry is a very complex centralized and vertically integrated system, roughly divided into four complementary sectors, the upstream, midstream, downstream and the service sectors; the industry includes many vertically integrated international and national companies. According to a social constructivist approach, the petroleum industry is an integration of technical, social, economic and political elements (Bijker et al., 1994), as any of the other complex systems can be pictured. Supplying about 60% of global energy consumption, the petroleum industry has endured a number of successive transformation stages. However, the industry as it now consists of numerous technological artefacts, social constructions and political interventions, that create a certain kind of ‘seamless web’ (Bijker et al., 1994)
– a mixture of society, technology and various prerequisites to either. Thus, as with any other complex system, the petroleum industry is a challenging object to study. Before we apply the Dahménian development blocks concept to the latest transformation period, we must identify those periods of transformation.

To define the relative border between the major periods of transformation in the petroleum industry we performed a literature review. There are certain identified approaches in the literature to study the dynamics of the petroleum industry – the Hotelling approach (the economics of exhaustible resources), the political approach, the cost approach (Baddour, 1997), the technological approach Smil (2010) and the historical perspective Podobnik (2006). Baddour (1997:144) argues that the periods of development of the petroleum industry are characterized by the evolution of marginal exploration-production costs. Chevalier (1973) (cited in Baddour, 1997:143), proposes two other factors in a related study, which are the cost evolution trend and the degree of social awareness. In his study, Podobnik (2006) uses a historical perspective on global energy, world energy consumption and production trends. He supplements this perspective with macroeconomic analysis, a special emphasis on crude oil and a transformation to a sustainable energy system. Some studies argue for either dominating social constructivism in the development of the petroleum industry (or of the energy industry as a whole) or technological determinism influencing the development path. Mitchell et al. (2012) make geopolitics the cornerstone of change in the petroleum industry. Smil (2010) analyses energy transitions from a more technological perspective. Another view of the petroleum industry development was proposed by the Energy Charter Secretariat (2007, 2011). This view is based on a market approach – it uses the oil price as a core component of the change in the industry.
The summary of the identified periods of the development of the petroleum industry is in Table 1.

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Cost based / Baddour (1997)</td>
<td>majors order (1929-73)</td>
<td>OPEC order (1974-86)</td>
<td>consumer-countries order (1986 to the present*)</td>
<td></td>
</tr>
</tbody>
</table>

* Present is defined according to Baddour (1997)
** Classification is based on the oil pricing mechanisms in the oil market

Table 2. Periods of the development of the petroleum industry

Each of the periods is, in fact, an interaction between societal needs, political and economic power and available technologies. The energy industry, in general, and the petroleum industry, in particular, is technologically intensive, which means that innovative technologies can substantially increase the production share and the quality of the products. It has also been strongly embedded into society and geopolitics through decades. At this juncture, we must stress the fact that those identified approaches directly correspond to the concept of development blocks and, in particular, the levels of structural tensions: technological, industrial and societal (Table 1).

The first clearly defined period begins in the aftermath of the Second World War. It is also the starting point of the analysis, as in this period, oil production experienced exponential growth, shifting rapidly from the dominance of coal (Figure 1). Natural gas production has risen steadily since that time. The later rise of the petroleum industry was initiated by eight major international oil companies, which held the market in a cartel-like system and were constantly expanding it. It was supported by growing post-war demand and economic restoration, in addition to anchored technological determinism in terms of the internal combustion engine. The dominance of the majors weakened in 1960s, when more companies obtained licenses to operate in North African and Middle Eastern oil fields. The period of OPEC’s hegemony had begun with this long transition and finalized with a number of geopolitical conflicts, two oil
shocks (1973–1974, 1978–1979) and discoveries of new resources outside of OPEC’s influence. Notably, some argue that 1946-1973 was the most stable operating oil market in history (1946-1973 years, Podobnik, 2006:96). However, relatively low oil prices led some of national and independent companies to establish OPEC in 1960 in contradistinction to the price-controlled order of majors. The 1970s brought changes in the actors’ respective positions (Baddour, 1997:149) and the market order, changing the development of the petroleum industry fundamentally. The competition between OPEC and the majors later transformed into an intra-OPEC competition, followed by huge crude oil price instability and its decreasing share in global energy consumption. By the end of the 1980s-1990, customers had begun to exert the most influence on energy markets in general and on the petroleum industry in particular. The next period of “consumer-countries order” was founded, according to Baddour (1997:154), by a price war triggered previously by OPEC. Relatively balanced demand and supply led to less variation in crude oil pricing. Additionally, new discoveries of hydrocarbon resources and the rise of natural gas consumption contributed to the shift towards more diversified and sustainable energy system. The next stage for the global petroleum market, and the energy market in general, came at the beginning of the 2000s. This stage might be characterized as an era of unconventionals and renewables. Mitchell et al. (2012:22) define this new era of the petroleum industry as a “simple model” of “new fundamentals” in which oil and natural gas compete with alternatives, such as shale oil and gas, deepwater and pre-salt deposits, and other unconventional resources. Many of these new supplies are highly dependent on high oil prices due to the relatively high cost of extraction, as economic slowdowns influence the transition to “a new era” of energy.

Those transformations in the global oil market were initialized and followed by technological, industrial and social changes. However, the crude oil price remains one of the main factors behind transformation pressures and bottlenecks for the petroleum industry and the global economy. This price and/or its volatility drives and/or stalls industrial development, redirects investment to other energy sources and/or opens development opportunities for fuel alternatives. Some of the previously discussed critical transformation periods were followed by or were the consequence of crude oil price changes, such as the 1973-1974 Arab embargo against countries supporting Israel and the Yom Kippur War, the peak period in 1980 due to the Iran-Iraq War, the drop in 1986 that was followed by Saudi Arabia increasing oil production share, and the 2001-2003 early-stage instability in the Middle East. During the 2003-2007 period of relatively stable oil prices, a number of structural tensions have occurred and later have resulted in variety of transformation pressures, such as the global economic crisis, technological innovations, climate concerns and geopolitical instability.
The petroleum industry has been preserved from radical structural changes for quite some time. However, no system is entirely closed. The petroleum industry is also linked with other global issues, such as climate change, economic growth, etc. Arguing that the petroleum industry development block has stepped into a ‘new era’ of disequilibrium from approximately 2008 onwards, it is reasonable to study what specifically initiated that transformation. We are now moving to an analysis of the tensions and pressures influencing the petroleum industry development block on its latest development stage.

**Tensions and pressures on the way to a renewed balance**

Currently, the petroleum industry development block is experiencing changes of all kinds and is highly imbalanced. Newly emerged pressures transform the petroleum industry development block and the overall disposition of complementarities. To investigate the latest transformation period, we apply the concepts of the development block, transformation pressures and structural tensions developed by Dahmén.

**Box 2: Transformation pressures, structural tensions and transformations**

*In 1989, Dahmén described the interaction of the development block elements, including transformation pressures and structural tensions. We would like to offer a short definition of the main elements once again, which are later supported with our analysis of the empirical case of petroleum industry.*

**Transformation pressures** are those pressures that influence the development block from the outside and force the system to adapt to the new operating conditions.

**Structural tensions** are multidirectional forces within the development block that occur when the system lacks balance and is being influenced by the transformation pressures. Those tensions occur because the system is actively seeking a new relative balance.

**Transformations** are those new elements appearing in the system, or ‘free space’ left by old elements, which are scrapped from the system. They are driven by innovations and new technologies.

To examine what the structural tensions and transformation pressures are, we use the “2015 World Energy Issues Monitor” by the World Energy Council (WEC, 2015). This particular study analyses the world energy agenda and its evolution over time. Based on the insights of 1045 energy leaders from 79 countries, the study highlights the list of the 40 most urgent world energy issues and evaluates their perceived impact and uncertainty for the energy

---

63 We use the WEC (2015) wording; by issues, we mean the transformation pressures, structural tensions and transformations combined.
sector globally. Those issues are mostly of new origin and initiate transformation of the industrial structure and ways of operating. Those issues have been further conceptualized for the petroleum industry, narrowed for purposes of this chapter and summarized in Appendix A.

The issues which arise on the technological level are technical advancements and also challenges that create the shifts in the petroleum industry, opening new transformation opportunities and filling the gaps in the development block. Technology and innovations continue to stimulate structural changes in the petroleum industry through rapid advancements in conventional and unconventional oil and natural gas exploration and development, as well as in alternative fuelling. The technical ability to extract unconventional resources (1.1) leads to the choice of which resource to develop, contributes to overall energy supply growth and its diversification, and spurs major structural changes in the energy markets. As long as consumption is growing and technological advances become available, oil and natural gas are being explored in remote geographies, difficult environments and from unconventional sources. This is the case with pre-salt deposits; resources located in deepwater, Arctic and other harsh environments; oil sands; shale; and other sources. These hard-to-access petroleum deposits have become economically viable for companies to explore because of high oil prices in 2011-2014. One of the cases of structural change involves shale gas and Arctic resources in the US. As the result of technological advances – and especially the shale revolution – the energy balance has shifted towards unconventional sources of oil and natural gas. New energy products, oil-avoiding technologies (1.2) in the face of liquefied natural gas (LNG), gas-to-liquids (GTL), biofuels, fuel and electric cells, and other unconventionally sourced energy resources compete on the energy market, where oil-avoiding technologies are actively searching options to replace oil products in transportation. The result of the overall competition is already visible in regional markets (Brazil and biofuels, LNG in Japan, etc.). Swift cost reductions in alternative energy will bring new insights to energy consumption. The energy efficiency of the petroleum industry (1.3) itself is increasing and allowing more energy resources for end use, which influences the competitiveness and ecological attractiveness of traditional petroleum products. The technological progress in energy efficiency, can make the concept of ‘peak oil’ weaker in the minds of consumers. Carbon capture and storage (CCS) (1.4) is a technology with a huge potential to mitigate climate change. This technology is capable of capturing up to 90% of the carbon dioxide emissions produced from the use of fossil fuels. Pilot projects have been implemented globally. However, the economic viability and public commitment for implementation of this technology is lagging. Large-scale accidents (1.5) influence the perceptions of traditional energy production and utilization, making society switch to alternatives. Examples of such dramatic events include the Exxon Valdez (1989) and the Deepwater Horizon
(2010) oil spill. Rising ecological concerns whipped up by such accidents create obstacles for the petroleum industry on the path toward quantitative development.

On the industrial level, there are a number of economic and market issues. They form global transformation trends and consist of structural, market, and regional changes and volatility in energy pricing. **Structural changes (2.1)** in the petroleum industry occur due to an additional competition among energy producers and energy sources themselves. The changing roles of net importers and net exporters of resources, or, in other words, changes the supply-demand geography, breaks established routines in the oil and natural gas markets. This is the case with shale gas in the U.S., where due to a ‘shale revolution’, the country can become more than self-sufficient in energy production, potentially changing from energy importer to energy exporter. Traditional supplies have been redirected to other energy markets, where they caused structural changes of the traditional energy balance (the case of Europe and cheap U.S. coal). Gas-to-gas competition and coal-to-gas competition are adding to structural changes, e.g., to the change in the global energy balance. **Regional changes (2.2)** consist mainly of shifting supplies from West to East, resulting in Asian-centred energy consumption. While European countries focus on reducing energy consumption in all spheres of life, other trends such as energy efficiency and the growing production of renewable energy and the consumption of energy resources in Asian countries is steadily growing. Additionally, there are changes in the Asian energy balance, where coal is being replaced by oil and natural gas. The greater role of China is appearing not only in consumption but also in acquiring oil and gas resources around the world to sustain its economic growth, i.e., its future energy consumption. Moreover, OPEC has switched export volumes to Asian markets and has caused small price wars there. **Market changes (2.3)** are expressed in global economic trends. For now, they include the global recession, decreasing GDP growth rates, margins and increasing concerns involving economic security. These trends have resulted in decreasing energy consumption growth rates and lower margins for energy companies, as well as increasing concern of energy security and supplies. Due to the overall recession, there is a growing challenge to mobilize investment capital into the energy infrastructure, energy resource development and other related projects. **Changing the role of OPEC (2.4)** becomes a reality under new conditions in the global economy and energy systems, where it is losing its power to dominate oil market prices and price formation by simply decreasing or increasing oil production volumes. Instability in traditional petroleum-producing regions, the diversification of supplies and energy resources, and new technologies have reduced the power of producers and given this power to the consumer. The actual power of consumers – the ability to choose and diversify – turned to be stronger than the power of having the resources. **The energy price and its volatility (2.5)** influence the overall layout of the
energy sector, and we consider them to be the most powerful driving factors at present. The energy price level affects new oil and natural gas resource development prospects, technological development, infrastructure, etc. The volatility of the energy price creates additional instability and unpredictability in the energy markets. Additionally, these forces add to the price imbalance among regional energy markets and between the energy resources themselves. The price of natural gas in Europe, the U.S. and Japan differs significantly, whereas price volatility expands the gap. The quality price margin between types of oil (WTI and Brent) has diminished.

The final level of analysis lies on the societal level and includes both geopolitical and ecological issues. Ecological issues play a critical role in contemporary changes in the industry, including corresponding policy and regulations, as well as generally increasing ecological public awareness. Geopolitical issues mainly influence the stability of oil and natural gas markets. Geopolitical instability and conflicts (3.1) loosen established connections and operational principles and result in, first, changing traditional supply-demand relations in energy markets. Traditional petroleum producing regions suffer from war, terrorism and other conflicts that affect oil and natural gas production and the energy infrastructure, thereby leading to concerns with the stability of energy supplies. This, in turn, affects energy markets and energy prices. Increasing awareness on the security of the supply (3.2) is again caused by the number of the previously mentioned drivers, such as energy price volatility, geopolitical instability, etc. The concerns result in an increasing urge towards securing the energy supplies for net energy importers (such as the EU and Asian countries). The instruments for gaining increased security are decentralized development in the energy system, smaller-scale energy projects and the diversification of supplies and energy resources. Government policies towards alternative energy and fuels (3.3) are one of the long-term instruments for stabilizing regional markets in terms of breaking oil dependency and creating a favourable investment climate for alternative energy through technological development, subsidies and tax regimes. This trend is developing further as alternative, unconventional and renewable energy sources become more available and less costly. The effect of this driver on the global energy system and the petroleum industry in particular is increasing as alternative energy resources gain greater share in the global energy balance. The climate framework (3.4) is forcing the energy system to change, bringing pressure to the petroleum industry. Society has become more sensitive to petroleum activities, first in relation to environmental protection and climate change issues. The Exxon Valdez spill and the development of the recent Macondo well in the Gulf of Mexico are two incidents that raised public awareness and generated more radical criticism of petroleum activities on the global and regional levels. These have resulted in greater attention to the overall mitigation of climate change, where demand for “green energy” and reducing CO2 emissions is growing. Reducing CO2
emissions (3.5) is one of the main instruments of the present day to fight climate change. Political incentives and the economic and technological ability to reduce CO$_2$ emissions regionally and globally are pushing for a further reduction in hydrocarbon energy resource consumption to achieve the desired goals.

![Diagram](image)

**Figure 4: The map of the petroleum industry development block transformation.**

We now map out all these issues according to the Dahménian development block concept; some are actual structural tensions, some are the transformation pressures, and most of the technological issues are those balancing the innovative gap-fillers/transformations. The development block of the petroleum industry is systematically shown in Figure 4.

The development block of the petroleum industry now has its shape and content, which gives us an opportunity to examine it in depth. As with any other system, the petroleum industry is being influenced from the outside (1.5, 2.5, 3.1, 3.3-5), and it reacts to these pressures with the structural tensions inside the system (2.1-4, 3.2) and produces gap-fillers to adjust to the new business reality (1.1-4).

The most influential transformation pressures according to WEC (2015) are the energy price, volatility and the climate framework. The energy price keeps the energy markets tense as it was decades ago. Economical and market issues still dominate and weigh upon the other non-market issues. It is also supported by market changes, which means decreasing GDP rates and
lowering consumption. Energy prices, specifically, crude oil prices, continue to play a leading role in the petroleum industry transformation. Its volatility is very much supported by geopolitical instabilities – another source of pressure in that development block. The non-resilience of the energy markets to energy price fluctuations has been demonstrated during the financial crisis of 2008-2009 and the latest dramatic decrease of the oil price in 2014-2015. As exemplified previously, the factors include the slow-down of shale, the coal supplies shifting from the U.S. to Europe, and ‘scuttle’ from the hard-to-access hydrocarbon resources. The energy price continually shapes the energy markets and is a major concern of global leaders (WEC 2015), companies and governmental institutions. The climate framework is the second-most influential transformation pressure currently. In general, transition to sustainability has now received huge interest in academia and politics (Verbong, Geels, 2007). Climate change mitigation has also invited a large substantial interest and has already exerted a large amount of pressure on the petroleum industry. As our analysis shows, environmental issues in the petroleum industry have overcome some of the cost and technological and diversification issues64.

The most important issues from the perspective of bringing balance to the system are technological innovations, the same as those defined by Dahmén decades ago. When we analyse the petroleum industry development block as shown in Figure 4, it is easy to see the ‘responses’ of the system to the outer pressures. They are the ability to extract unconventional and new energy (oil-avoiding) products, CCS and energy efficiency. Those issues, to some extent, stabilize the development block and open new opportunities for innovation. In a period of high oil prices, costly Arctic oil and natural gas resources appeared to be a good complement to other hydrocarbons from the more traditional energy producing regions. New energy products (and forms of energy) are being constantly sought and developed. One example is gas hydrates (actively researched by Japan, a “frozen” mixture of natural gas and water under the sea bottom and/or high pressure). A widespread form of a new natural gas product is liquefied natural gas (LNG). LNG is traded all over the world but first appeared in 1960s (Mckenzie, 2012). Some of those new types or forms of energy resources limited the market shares of the older ones, such as compressed gas and fuel oil. The special aspect of the petroleum industry development block is that none of the older oil or natural gas products are completely scrapped from the market. Driven by the climate factor, the search is directed towards oil-avoiding technologies, e.g., biofuels. However, at the time of this writing, those technological innovations (gap-

64 Verbong and Geels (2007) argue that environmental issues in energy systems follow cost, reliability and diversification issues, where the transition is more driven by liberalization and Europeanisation.
fillers) also create structural tensions. New products and unconventional resources in face of shale have destabilized the global markets and created an oil and natural gas overflow. The ability to extract distant and hard-to-access resources has created a misleading path for the capital investments. However, it is worth mentioning that most of those technological innovations have been created inside the system and are the responses of the system to the outer transformation pressures and inner structural tensions.

To summarize, the current state of the petroleum industry development block is clearly in a disequilibrium stage. Numerous tensions and pressures push the system toward further transformation. One of the reasons for this is the incoherent transformation of the complementarities. However, the industry development is not a linear process – it can grow, stagnate, step in the period of tough competition or fade away. New technological innovations can radically transform the system or may flow in a predetermined technological pattern. Because the petroleum industry is an extremely large technoeconomical system with tight connections to society, it also possesses a certain level of resilience. On the one hand, it is hard to escape a certain development lock-in when the system is relatively large. On the other hand, the actual development stage and prerequisites for the system development and transformation, such as financial resources, are important to facilitate the costly transformation. A majority of tensions and pressures have been shaping petroleum industry for decades and will do so for decades to come. However, the relatively new transformation pressures, such as climate change issues, are increasing in weight and strength. These new pressures are pushing the petroleum industry out of the existing deterministic path.

Conclusions, managerial and policy implications

The energy sector is one of the largest in the world economy. The economic and social weight of energy, especially in the petroleum sector, is huge. The growth of the energy sector in general and the petroleum sector in particular during the last century is dramatic. It has gone through stages when wood and coal were the major energy resources and can be found on a stage where oil and natural gas dominate. In this chapter, we suggested looking at the transformation of the petroleum industry through the prism of the development block concept of Dahmén.

The periods of transformation of the petroleum industry can be identified with the help of the key transformations on technological, industrial and social levels. We addressed several studies on the petroleum industry development and analysed them based on the Dahménian approach. The analysis of the structural tensions and transformation pressures of the petroleum industry development block in its latest period has been built on the study of the WEC
The defined 15 petroleum industry issues are the areas of economics, geopolitics, technology and ecology. They have also been grouped according to the three levels of tension by Dahmén: technological, industrial and societal. Later, the structural tensions, transformation pressures and the gap-filling innovations were defined. The two transformation pressures have a higher impact on the petroleum industry transformation – energy price and volatility and the climate framework. The technological innovations are the stabilizing elements for the development block, but they also create structural tensions within the system.

The analysis shows the complexity of the petroleum industry. It describes its industrial development, which is a non-linear process, and its current state on the path to the next relative balance. Although the majority of those tensions and pressures are well known and studied, their impact on the whole system is not researched. The relatively new pressures, such as climate change issues, have more importance currently compared to before. These tensions and pressures can be of interest to other studies, especially in the field of industrial dynamics.

This chapter has shown one of the approaches within the field of industrial dynamics to study complex industrial/technological systems. One of the main advantages of the development block concept is that it uses a systematic approach and allows the use of other theoretical frameworks within the systematic analysis. Another advantage is that it offers the possibility of combining several layers of analysis at the technological, industrial and societal levels. Third, the development block concept is a platform to provide material for empirical investigations in economic-historical research, where it also offers an opportunity to contribute to the theoretical discourse. Finally, the development block concept is a good analytical framework to conduct multidisciplinary studies. We also explicitly underline the importance of the interdisciplinary view in those types of studies. We see the development block concept as a useful tool for a system analysis in various industrial systems, such as the energy sector, metallurgy, other “heavy” producing industries, and also those with a more innovative character like ICT.

In the case of the petroleum industry, the applied system approach of the Dahménian development blocks can lead us to even more interesting discussions, which are some of the managerial and policy implications of the research. Some further and deeper structural tensions of the petroleum industry are capital divestment from oil exploration and infrastructure projects and oil-to-natural gas substitution. Capital from the richest funds – pension funds – starts to move away from oil-based long and medium term projects given the pressure of climate change mitigation (IBTimes, 2015; Reuters, 2015). These tensions can be characterized as market or regional tensions. However, they have a broader and multidirectional influence on the petroleum industry, energy-developing and producing companies, and
service companies. This has a negative influence on hydrocarbon reserve base renewal: the growing need for long-term investments into more expensive resources exploration is not well supported. This forces energy companies to search for alternatives, whether it is enhanced oil recovery in the ‘brown fields’ (those being developed for period of time) or extensive development of the existing reserves. Producers and mostly consumers must secure energy supplies, which means ensuring the good condition of the energy infrastructure. Another tension arises in the urge to decrease CO2 emissions. Natural gas replaces oil in many processes because it produces far fewer emissions. Oil has become more of a ‘niche’ product than in previous decades, taking the highest shares in transportation and petro chemistry. The forecasted growth rates for natural gas are double that of oil (1.8% per year for natural gas to 0.9 per year for oil, BP, 2016a). These projected changes assume a re-forwarding of the investments and the reshaping of some businesses (e.g., power production plants, heavy industries, etc.). Since the overall trend of oil consumption is expected to decline in the near future, it will influence the petroleum industry and require corresponding governmental and managerial actions.

References


Appendix A: Petroleum industry issues

The issues are grouped according to the three levels of the structural tensions – technological, industrial and societal (Table 1).

Table A1. Petroleum industry issues (compiled by authors based on WEC, 2015:9-10)

<table>
<thead>
<tr>
<th>1. TECHNOLOGY</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>1.1. Technical ability to extract unconventional resources</td>
<td>Changing the energy balance towards unconventional energy resources due to their increasing technical availability</td>
</tr>
<tr>
<td>1.2. New energy products, oil-avoiding technologies</td>
<td>New products (LNG, biofuels, etc.) and oil-avoiding technologies (fuel cells, electric cells, PV, electricity storage) compete for a higher share in energy markets and overcoming barriers to realize their potential</td>
</tr>
<tr>
<td>1.3. Energy efficiency of the petroleum industry</td>
<td>Energy efficiency of the industry itself to achieve higher potential of utilized fossil fuels, allowing more energy resources for end use</td>
</tr>
<tr>
<td>1.4. Carbon capture and storage (CCS)</td>
<td>Expansion of CCS technologies to overcome the existing barriers to make CCS viable with regard to cost</td>
</tr>
<tr>
<td>1.5. Large-scale accidents</td>
<td>Critical large-scale accidents influencing the perceptions of traditional energy production and utilization</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>2. INDUSTRY</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>2.1. Structural changes</td>
<td>Changes within the structure of the global and regional markets, including interconnection, changing supply-demand geography, and energy resource competition</td>
</tr>
<tr>
<td>2.2. Regional changes</td>
<td>Asian-centered energy consumption, demand and supply</td>
</tr>
<tr>
<td>2.3. Market changes</td>
<td>Global recession, decreasing growth rates, margins and economic security. Challenge to mobilize investment capital to energy projects</td>
</tr>
<tr>
<td>2.4. Changing role of OPEC</td>
<td>OPEC is losing its trigger role in global oil market and price formation</td>
</tr>
<tr>
<td>2.5. Energy price and volatility</td>
<td>Global oil price change and volatility, price ratio gaps for energy resources, technologies and regional price differences</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>3. SOCIETY</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>3.1. Geopolitical instability and conflicts</td>
<td>Geopolitical instability and increasing vulnerability of the supply from traditional petroleum-producing regions, terrorism, other physical risks affecting energy system, infrastructure and markets</td>
</tr>
<tr>
<td>3.2. Increasing awareness of the security of supply</td>
<td>Security of supplies as a trend in energy markets, a tendency towards decentralized energy systems, smaller-scale projects and diversification of supplies and energy resources</td>
</tr>
<tr>
<td>3.3. Government policies towards alternative energy and fuels</td>
<td>Favorable investment climate for alternative energy development and technology implementation, subsidies for alternative energy and fuels, higher taxes on fossils</td>
</tr>
<tr>
<td>3.4. Climate framework</td>
<td>Global climate mitigation, policy and regulations at the global and regional levels, increasing public awareness on ecological issues</td>
</tr>
<tr>
<td>3.5. Reducing CO₂ emissions</td>
<td>Political incentives, economic and technological ability to attain CO₂ emission reductions globally and regionally</td>
</tr>
</tbody>
</table>
Appendix B: Energy data sources

Data for Fig. 1 – Global energy consumption shifts on biomass, coal, crude oil, natural gas, hydroelectric power, nuclear energy and renewable energy (geothermal, solar, wind energy and other sources, excluding biofuels) were acquired from the following sources (applicable for all types of the previously mentioned energy sources if the contrary is not mentioned): 1800-1960 Smil (2010:Appendix), data available for each decade; 1965-1990 BP (2015c) on a yearly basis and 1990-2003 for biofuel production BP (2015c); 1991-2001 BP (2002); 2002-2003 BP (2006); 2004-2014 BP (2015a) including data on biofuel production and renewable energy production; 2015 BP (2016c); forecast data for 2020-2035 on a 5-year basis were sourced from BP (2015b), including biofuel consumption.

The data on biomass energy consumption is based on Smil (2010:Appendix), where the random number generation between the data on previous and the next decade was used to fill the data gaps for the period 1965-2014. It is also applicable for the period 2009-2014. Volumes of biofuel production (1965-2014) and biofuel consumption (2015-2035) were included in the total energy consumption from biomass and excluded from renewable energy consumption. In the period 2015-2035, the average percent of energy consumption from biomass from total energy consumption was fixed at 7% plus biofuels.

Data on hydroelectricity consumption for the period 1965-2014 was acquired from BP sources (2002; 2006; 2015c) and reversely extrapolated to the period 1880-1960 with an average year growth rate of 3% (which is an average growth rate of hydroelectricity energy consumption in 1966-2014).

The idea of the figure was developed from Podobnik (2006:5).
Introduction: why is catching up important to understand?

Technological catch-up by industrial latecomers is possible. This was demonstrated a century ago by Thorstein Veblen (1915), a Norwegian–American economist, in his study on Germany’s transformation from economic backwardness (in the 18th century) to a leading industrial economy.

Looking further back in history, the 17th-century technological leader, Holland—having boomed in the so-called ‘Golden Age’—lost its edge to Great Britain (Mokyr, 2000); Great Britain, in its turn, got caught up by continental Europe (e.g. Germany) and by America in the 19th century, in level of industrialization (Landes, 1969).

Looking closer in history, Japan is another example illustrating a technological catching up, prior to (and during) the Second World War period (Freeman and Soete, 1997). Following in the footsteps of Japan, the so-called ‘Asian tigers’, namely Singapore, South Korea, Taiwan and Hong Kong joined what Baumol et al. (1989) called ‘convergence clubs’ in the Post-Second World War period. ‘Convergence clubs’ (Baumol et al., 1989) mainly refers to the fact that a technological catching up—and consequently an economic convergence towards the frontier—seems to be only confined to groups of countries, rather than applying to all (late-coming countries).

More recently, the rapid technological upgrading and industrialization of China, and perhaps soon also of India, provides additional illustrations, and shows that technological catching up by latecomers is indeed possible.

Technological catch-up by industrial latecomers occurs also at an industry and a firm level:
William Henry Perkin (1838–1907), an English chemist, at the age of 18 accidentally discovered the first aniline dye, mauveine, and quickly found a way to produce it. The synthetic dye industry was therefore born in 1856 in England. While the dye quickly became fashionable in many other countries, like France and the US, many expected that Britain would remain as the leader of synthetic dye manufacture. This leadership, however, only lasted for eight years, according to Murmann (2003): German firms quickly learned molecular combinations for popular fabric dyes and managed to catch up to, and forge ahead of, their English competitors, in the late 19th and early 20th centuries in terms of technological advancement.

A hundred years later, the same phenomenon also hit the liquid crystal display technologies (Murmann, 2003): it was demonstrated in the US in 1968; produced in Japan in high volume in the 1990s; led by South Korean manufacturers at the beginning of the millennium and currently is mostly developed in Taiwan and mainland China (in important sub-categories).

More recently and at the firm level: at the fifth-generation (5G) mobile system technologies’ front, Huawei—a China-based and less than 30-year old firm—races against Ericsson, who has dominated the field of communication equipment supply for more than 100 years. At the service provision, rather than the product supply-front, Alipay—owned by Alibaba and based in China—is catching up to the US-based PayPal, as a third-party-based payment (service) platform. Already by 2014, Alipay gained more than 300 million users (against PayPal with about 250 million users).

Questions: why have some countries become industrialized and got rich while others stay poor? Why have some industry sectors taken the lead in their nation’s industrialization processes, while others remain as followers? Why have some firms, despite a late entry, managed to catch-up and forge ahead, while others, including the earlier incumbent ones, fall behind?

More importantly, what are the causes and mechanisms underlying such a dynamic shift of the leaders’ versus the laggards’ positions?

The aim of this chapter is therefore to explore:

- How has industrial catching up as an empirical phenomenon been hitherto explained;
- How could the literature be used in analysing modern cases?

This chapter is structured as follows: Part I offers a scholarly definition on industrial catching up; Part II reviews some of the propositions discussed in the literature; Part III goes further to discuss the analysability of the theories and provides case illustrations. At the end, managerial and policy implications are briefly discussed.
**The concept**

Industrial catching up is often discussed at three different but interconnected levels: macro (national), meso (sectoral) and micro (firm) levels.

*The macro (national) level definition* on industrial catching up is often associated with two terms: convergence (versus divergence) and growth.

“Catch-up’ relates to the ability of a single country to narrow the gap in productivity and income vis-à-vis a leader country, ‘convergence’ to a trend towards a reduction of the overall differences in productivity and income in the world as a whole.”

Fagerberg & Godinho (2005, p. 514)

The other side of the ‘convergence’ coin is ‘divergence’: 250 years ago the ratio between the richest and the poorest countries was only 5:1 and now it is 400:1, according to a famous economic historian, David Landes (1998). This is based on the measurement of income and productivity. The very existence of the ‘convergence club’ members (Baumol et al., 1989), for example of Asian ‘tigers’, is, therefore, a myth against the general divergence background and worthy of exploration.

Macro level studies are often prone to a search of input versus output indicators, for example productivity levels (through examination of gross domestic product (GDP)/capita data). This brings us to another often-used concept: growth.

Catching up, if persistent for a longer period, often leads to a process of macroeconomic growth, and consequently of industrialization. It is not uncommon to see that the three concepts go hand-in-hand. For example the history of Swedish industrialization is also a process of (industrial) catching up to that of England and continental Europe (Bruland, 1991). Looking closer in history, the Chinese catching up now is also synonymous to its industrialization (accompanied by a rapid urbanization).

Nevertheless, it is important to keep in mind the conceptual differences and be aware of the causalities involved (e.g. the ‘chicken and egg’ problem).

Catching up often implies also (technological) capabilities’ enhancement, which in turn leads to increase in productivity and growth. Macroeconomic growth—often using GDP as indicator—however, does not have to be linked with industrial catching up. For example the recent miraculous sub-Saharan Africa (SSA) growth is mostly attributed to the boom in commodity prices (e.g. minerals and like natural resources). This kind of growth is hardly linked with (capabilities-based) catching up, though arguably agro-processing sector productivity increase is also part of the growth story.
The meso (sectoral) level definition of industrial catching up often focuses on sectoral-specific dimensions: a) Knowledge, as well as technological opportunities, differs across industry sectors (Malerba and Orsenigo, 1996). That, in turn, affects the appropriation means used by industry actors. Appropriation refers to ‘profiting from innovation’ in Teece’s (1986) terminology. This sectoral pattern of knowledge generation and knowledge appropriation in its turn, affects the method and strategies used by late-coming firms in catching up (Long, 2014); b) Institutions and industry structures also differ across industries, constituting sectoral patterns of industrial catching up (Malerba and Nelson, 2012) and c) Mechanisms like co-evolution between industry structure, institution and technologies in industrial catching up are also to be addressed (Murmann, 2003; Nelson, 1995).

Measurements at the meso level—on industrial catching up—are close to those on innovation, namely from the output side on new products, new processes, patents, industry turnovers, and from the input side on R&D investment, R&D intensity, etc.

The micro (firm) level definition on industrial catching up often has a (technological) strategy in focus. Innovation is relevant here in the sense that firms’ catching up is tightly linked with the issue of technological trajectory, namely a choice of technological path(s). Measurements at the micro level—on industrial catching up—are close to those on firm competitiveness, for example market share and dominance in (technological) standards, etc.

One often-raised question (on industrial catching up) at the meso and micro levels is: who is catching up with whom? Is Apple considered a latecomer in mobile communication (through producing a smartphone after 2007) and catching up to (or even outstripping) Nokia?

The answer is ‘no’. Apple’s ‘intrusion’ into mobile communication—considering its origin in data communication—is more an issue of technological convergence, rather than an issue of catching up.

To clarify, the catching-up firm often is:

a) Initially lagging behind, in levels of technological sophistication, compared with that of the world leaders;

b) Often based on late-coming developing countries, with a locational disadvantage of not being physically present in, or adjacent to, world-class technology centres. The latecomer is therefore not, by definition, able to enjoy the Marshallian knowledge spillovers ‘knowledge (existed) in the air’;
c) Capable of catching up (alternatively forging ahead) in technological sophistication.

The technological catching up of the Chinese semiconductor industry (Ernst, 2015) may illustrate this definition (while the world frontier is in Silicon Valley).

**The theories**

What explains industrial catching up? In the following, it is not my intention to repeat the standard textbook messages and list the factors having a bearing on catching up. Rather, it is a review on key propositions related to that.

**Factors**

The key propositions related to catching-up factors may embrace:

a) *Latecomers carry a theoretical potential to catch up*

Moses Abramovitz (1986; 1994), an economic historian, argues that productivity growth rates tend to vary inversely with productivity levels. So, the further behind a country lies, the larger the potential it has to catch up. In other words, being backward carries a theoretical potential.

This argument may date back to Gerschenkron’s (1962) discussion on the ‘advantages of backwardness’ and Veblen’s (1915) ‘disadvantages of technological leadership’ (in the case of the then imperial Germany).

What Abramovitz (1986) has extended in the 1980s is the ‘social capabilities’, a concept conveying that factors like education levels in the society are important conditions to realize that potential (a qualification of the catching-up hypothesis).

b) *Catching up is an accumulation of factors*

The classic accumulation school of thought (on catching up) argues that industrialization, in general, is a process of capital accumulation, or of accumulation in (general) factors. Catching up here is treated as a subset of industrialization.

The massive investment in physical and human capital made by Asian newly industrialized countries (NICs) for example is argued to be a sufficient explanation of the (then) ‘Asian miracles’ (in catching up) (cf. Kim and Lau, 1994; Krugman, 1994; Rodrik, 1995), and also in China’s case today (World Bank, 2014).
This (classic) accumulation view may date back to Karl Marx’s (1867) economic theory on surplus value (reinvesting profit into the economy), to Harrod–Domar’s (Harrod 1939, Domar 1957) model on savings-pushed growth (that can outperform population growth) and to Robert Solow’s (1957) addition on labour being (also) a factor included in neoclassical growth accounting.

An opposite—or perhaps a complementary—school of thought is the so-called ‘assimilation’ approach, which addresses the role of learning, innovation and entrepreneurship in industrial catching up (to be presented in Section 2.2 on catching-up conditions).

c) Institutions matter in catching up

Institutions matter in economic growth (Acemoglu et al., 2004) and in development (e.g. industrial catching up) (Rodrik and Subramanian, 2003). Whether this factor, namely, the quality of institutions, overrides other factors (e.g. geography, trade) in explanation or not, however, remains an issue of long-standing debate. Jeffrey Sach (2003), for example argues that institutions matter, but not for everything.

Fagerberg and Srholec (2009) suggest a distinction between different kinds of institutions to assess the role of the institution in catching up: institutions in North’s (1990) sense, namely those slowly evolved into ‘rules of the game’ (e.g. judicial system) and institutions reflecting political choices, such as policies.

The first category of institutions—the Northian sort—can be exemplified with international ones such as the World Trade Organization (WTO), Trade-Related Aspects of Intellectual Property Rights (TRIPS) and national ones such as the national Intellectual Property Rights (IPRs) laws. They are not good predictors of successful catching up.

The second category of institutions—for example industrial policies—is often subjected to frequent change. They are better predictors of catching up compared with the first ones; the debate on this category of the institution is often associated with the discussion on the role of the state in capitalist development, as follows.
d) The role of the state in catching up

The rationales behind the role of the state have always been highly controversial. The different positions may date back to Adam Smith’s (1776) ‘invisible hand’ argument, advocating the power of the free market in promoting competition and hence also in allocating (scarce) resources.

Contemporary debates, exemplifying the ‘liberal market economies’ (LMEs) (e.g. US, UK) versus the ‘co-ordinated market economies’ (CMEs) (e.g. Germany, Japan and Sweden), discuss also the different forms of the state’s involvement, under the umbrella of the Varieties of Capitalism (VoC) framework (Hall and Soskice, 2001). Similar arguments were put forward in the ‘three welfare-state regimes’ (Esping-Andersen, 1990). Policy terms such as the ‘Washington Consensus’ (Williamson, 1993) suggest market-oriented policies also in developing countries (cf. Gore, 2000).

The opposite arguments, namely those for the ‘visible hand’ of the state arguments, are also numerous and may date back to:

a) Alexander Hamilton’s argument (1791) for infant industry protection (to enable the then new United States to catch up with the United Kingdom);

b) Friederich List’s (1841) discussion of the role of policy and institutions as regard Germany’s need to catch up (to Britain);

c) Gerschenkron’s (1962) historical essay describing that (e.g. financial) institutions and industrial policies are instrumental for Continental Europe’s catch-up (also) to Britain in the 19th century.

This is also the case of industrialized countries. For example the ‘entrepreneurial state’ (Mazzucato, 2013) and similar concepts have rejuvenated the state-led policies arguments in discussing for example renewable-energy-related industry sectors.

The classic rationales of the above-mentioned state-led policies (sometimes also called ‘visible hand’ or ‘interventionist approach’) include co-ordination failures in markets, information deficiencies hindering for example the creation of externalities. Arguably, these problems get aggravated in developing countries: ‘technological latecomers countries suffer from more pervasive market failure than earlier industrialized countries’ (Altenburg, 2013; p. 346).

e) Geography matters for catching up

There are at least three waves of ‘geography matters’ debates that are related to catching up:

The first wave of the debate may trace back to the Dutch Disease concept addressing the extent to which natural resources have played a role in
economic development. *Dutch Disease*, coined in 1977 by *The Economist*, was used to describe that the discovery of a large natural gas field in the Netherlands in 1959 actually deterred the development of Dutch manufacturing sectors. Corden and Neary (1982) have theoretically advanced the *Dutch Disease* concept in relation to an analysis on the pros and cons of natural resources' extraction in an economy. The basic prediction of the *Dutch Disease* concept is that a climate/geography conducive to rich raw materials reserves (e.g. Saudi Arabia's reserve of oil) can affect a catching-up economy in a negative way through bringing in inertia in structural change. Angola for example can survive with its minimal reserves (supplying to e.g. China) without having to build up industrial capabilities today, a phenomenon often also called 'resource curse' in the literature (cf. Sachs and Warner, 1995). Then the case of Norway and its industrial capabilities development today may show that being rich in natural resources can also be a blessing (rather than a curse). More conventional is the opposite argument, namely the so-called ‘developmental constraints’ (originally a biological term; Smith et al., 1985) often also including lack of natural resources. This is exemplified with the cases of Taiwan and South Korea (Shin, 1996), which were ‘forced’ to build up industrial capabilities as there were no natural resources to exploit.

The second wave of the debate may go to Jeffery Sachs’ (2003) emphasis on the role of disease in affecting economic development. In his elaboration, SSA is (still) trapped with AIDS, tuberculosis and malaria. ‘It had a climate conducive to year-round transmission of malaria and was home to a species of mosquito ideally suited to transmitting malaria from person to person’ (ibid., p. 39). The outbreak of a disease, as a factor in isolation or in combination with an unfavourable geography (e.g. SSA; highlands of central America), can also lower the Foreign Direct Investments (FDIs)’ inflow dramatically. To summarize Sachs’ argument: the ability of a disease to cut off economic development cannot be underestimated; an unfavourable geography can aggravate that plight by raising the cost of international trade.

The third wave of the debate is connected to innovation studies addressing terms like ‘externalities’ and ‘spillovers’. Much of that debate originated from the Marshallian (1920) ‘industrial atmosphere’ argument, sometimes referred to as ‘knowledge in the air’ (ibid.), that is, knowledge existing in the local firms and industrial activities. In the learning and catching-up context, technology diffusion is much about capturing knowledge spillovers embedded in that process. Empirical studies from this perspective include the Chilean wine industry catching-up case (Giuliani and Bell, 2005).

Independent of—as well as related to—the above three waves of ‘geography matters’ debates, is the logic of agglomeration economies (core versus periphery) that applies also to industrial catch-up.
It dates back to the ‘Circular and cumulative causation’ theories developed in the 1960s. It addresses polarization between the Advanced Industrial Countries (AICs) and the less-developed countries (LDCs), advanced by Nobel Laureate Gunnar Myrdal (1968), a Swedish economist, on the basis of an ‘Asian drama’ of overpopulation. In this line of argument, vicious circles apply: technology and capital get sucked increasingly into the centre and away from the periphery; poor countries became ever poorer.

To connect the theories, this logic goes into the fundamentals of *path dependence* and *lock-in* discussed by Arthur (1989) in evolutionary economies. ‘Previous capital is needed to produce new capital, previous knowledge is needed to absorb new knowledge, skills must be available to acquire new skills and a certain level of development is required to create the infrastructure ...’ (Perez and Soete, 1988, p. 459).

**Question:** What on earth can break down the vicious circle mentioned above?

As I see it, the role of (international) immigrants—with their impact on knowledge spillover—may be reassessed here. The world has never seen such a big scale of cross-border flows of people in general and of young students in particular. According to China’s Ministry of Education, nearly 500,000 Chinese students went abroad in 2014 (an 11% increase over 2013; 92% of that is self-funded) and nearly 400,000 students returned to China in the same year. According to the same source, it is estimated that 3.5 million Chinese have studied abroad (since the reform in 1978) and the total return rate is 75%.

This is Chinese data alone, not to mention India and other emerging economies. Saxenian (2006) documented the ‘brain drain’ versus ‘brain gain’ of Indian and Chinese overseas returnees (in her so-called ‘diaspora effect’) and the empirical world has progressed a step further since then.

The phenomenon of returnees is not new in the industrialization process, such as America’s documented catching up to the UK being benefited with engineers returned to the US from UK (Landes, 1969), and also the case of Nordic countries (Bruland, 1991). The difference today is the scale. To what extent this scale is modifying the catching-up parameters is worth going further in studies.

**Conditions**

In the macroeconomic literature, catching-up conditions—embracing histories, policies and even doses of luck—are traditionally treated as exogenous factors. In contrast, historians often incorporate these conditions as endogenous factors and discuss them in a path-dependent manner, namely arguing that ‘sleeping’ conditions can be activated with indeterminate consequences.
What are the conditions still applicable to today’s industrial catching up of for example Vietnam and/or SSA countries?

Here, I would like to distinguish the ‘hard’ and ‘soft’ conditions and go in depth into a particular ‘soft’ condition—capabilities—and introduce the development of the assimilation school of thought (see also Section 2.1). ‘Hard’ conditions here refer to physical conditions such as infrastructures as well as (intellectual) capital, land and labour reserves.

(Technological) capabilities as an argument addressed in industrial catching up may date back to the mid-1980s. Arguably, it starts from Nobel Laureate Amartya Sen’s (1985) abstraction ‘commodities and capabilities’, whereas capabilities is THE condition for anyone to perceive and consequently capture the opportunity. In the 1990s, the United Nations Development Programme’s (UNDP, 1990)—a programme led by Sen—capabilities were explicitly set as an important research agenda.

It has further been advanced along with the development of the assimilation school of thought (Nelson and Pack, 1999) on catching up after the 1980s. On the one hand, the successful industrialization of Japan and the NICs—countries capable of accumulating capital and assimilating technologies—suggests a more optimistic view of the ‘Asian drama’. On the other hand, it was identified—from the experience of Asian ‘tigers’—that learning and assimilation were far from automatic and required an extensive commitment to implementing and developing technologies. In other words, accumulation of physical and human capital only is not sufficient to explain catching up. Arduous learning, risk-taking entrepreneurship and innovation are to be taken into account in the catching-up process (cf. Freeman, 1994; Kim 1997).

Nelson and Pack (1999) name this ‘assimilation theory’ (on catching up), versus the classic ‘accumulation theory’ mentioned in Section 2.1. These two are not treated as opposite but complementary approaches.

Parallel to the Asian ‘tiger’ studies, Katz and his colleagues, in their influential work (Katz et al., 1987) studying Latin American manufacturing firms, address the importance of successive stages of learning: learning to operate, learning to invest, learning to innovate. This kind of learning, as argued by Katz et al. (1987), is particularly important for domestic firms after having imported foreign technologies.

Katz et al.’s (1987) work has nourished thinking on the importance of capabilities, accumulation and learning for development and catching up, illustrated with Lall’s (1992) and Bell and Pavitt (1995) studies.

Moreover, the ‘technological capabilities’ arguments also address the importance of process technologies rather than new products (Viott, 2002).
This departs a wider view of technology than R&D only. That is to say, capabilities are embedded in process technologies too (not only in product technologies).

The term ‘capabilities’ comes in various forms and is not only addressed in a development context; to exemplify:

a) Abramovitz’s (1986) famous ‘social capabilities’ concept (measured by e.g. country’s educational levels) is discussed as an important qualification of catching up;
b) Cohen and Levinthal’s (1990) ‘absorptive capacity’: A term describing ‘the ability of a firm to recognize the value of new, external information, assimilate it and apply it to commercial ends’ (Cohen and Levinthal, 1990, p. 128);

**Paths and trajectories**

Industrial catching up is also related to the choices of paths and trajectories:

a) **Innovation versus imitation**

There is a long-standing debate: is matured product/technology a default point of entry for latecomers? Or should the latecomers start from something totally new? The former indicates an imitation path and the latter addresses innovation. The former is related to learning through ‘ready-made’ technologies (Veblen, 1915) and the latter argues to ‘jump directly to the (technology) frontier’ (Gerschenkron, 1951; 1962). The former is included in the theories of ‘dependency’ (also on technology) and the latter incorporates diffusion study and argues that knowledge diffusion, by nature, embraces a certain degree of innovation (or innovative adaptation).

Thorstein Veblen (1915) argues that latecomers could take advantage of the existing ‘ready-made’ technologies to catch up, and that includes also the so-called ‘machine technology’, which is process technology in today’s terms. Implicitly, successful technology diffusion is sufficient. Without having to share the costs of its development, this—that is exploiting existing technologies—can be a very profitable affair (in catching up) (ibid., p. 249).

Veblen is not alone in this position. Today, many for example pro-globalization researchers also ask: if there is already a well-tested prototype available (in some other part of the world), why should a follower reinvent the wheel? Implicitly, innovation is not needed. Diffusion, per se, is sufficient when the division of labour is globalized.
Gerschenkron (1951; 1962) on the other hand, argues that firms should jump directly to the (technology) frontier and learn from the most advanced technologies. Reading between the lines, Long (2014) argues that Gerschenkron (1962) still just refers to adoption, rather than innovation. While learning by nature is cumulative, there is no shortcut to innovation, that is capabilities take time to build up.

The typical modern illustration of Gerschenkron’s ‘jump directly to the (technology) frontier’ may be that the SSA regions—without any existing infrastructure—can deploy state-of-the-art photovoltaic technologies and (potentially) become ‘greener’ than the developed regions. Morocco, at the time when I write this chapter (year 2016), is on the way to building the world’s largest concentrated solar power plant, which would be powered by the Saharan sun (that is stable and with big coverage). The ambition is to provide half of the country’s energy by 2020 (according to The Guardian 26th Oct 2015).

The fundamentals of the Gerschenkron catching-up path do not differ from the Veblen (1915) path, as I see it. Adoption is not the same as innovation. Entering the race close to the technology frontier requires a step further, namely from learning to innovation.

Innovation by latecomers is still considered not justified, mostly from the resource constraints (for input) perspective. ‘This is a potential that reflects these countries’ greater opportunity to advance by borrowing and adapting the best practice technology and organization of more productive economies’ (Abramovitz, 1994b, p. 87). The Schumpeterian innovation is, therefore, by default, considered the privilege of the developed world and imitation (e.g. technology transfer) is treated as THE path for learning and catching up.

Questions remain and are to be phrased in different ways: Is imitation the most plausible path? Is imitation sufficient? Is innovation necessary?

Perez and Soete (1988) argue that ‘mature products are precisely those that have exhausted their technological dynamism’ (p. 459) and this path indicates a fixed low wage and low growth development pattern. The ideal position, as Perez and Soete (1988) argue, would be an early imitator or an innovator and there are entry costs as well as windows of opportunities for that.

How about the so-called ‘IPR thicket’ (Shapiro, 2001) that has functioned as a pronounced new-entry barrier in pro-patenting sectors such as information and communication technology (ICT) and biotech industries? The ‘IPR thicket’ (Shapiro, 2001) refers to a dense web of interconnected patents that are built upon each other (but often owned by different firms). This is a new barrier that has primarily emerged since the 1990s. How could the latecomers,
particularly those in the pro-patenting industry sectors, manage to navigate this thicket?

Long (2014) illustrates this with the Chinese ICT firms (e.g. Huawei) catching-up example. The results reveal that patent application and domestic technology standards' creation go hand-in-hand in this catching-up process. In other words, innovation, particularly the cumulative kind, is needed. There is a strong correlation between industrial catching up and innovation. This view is shared by Kim’s (1997) study on South Korean industries’ case ‘from imitation to innovation’.

b) ‘Flying geese’ versus ‘leapfrogging’

Development economist Akamatsu (1935; 1962) addresses the Japanese ‘flying geese’ model, namely one goose leads the flying (catching up) while the rest (geese) follow, in a structured and gradually progressing way.

‘Leapfrogging’, in contrast, refers to stage-skipping catching up. The Irish ICT (mostly software industry)–towards the end of the 1990s–is one of the typical cases of Irish upgrading from the traditional, low-skilled (or medium-tech) activities to one of the most ‘high-tech’ economies in the world today (Fagerberg & Godinho, 2005; O’Sullivan 2002). There are similar cases with regard to the Israeli software industry (Arora and Gambardella, 2005).

c) ‘Big scale’ versus ‘small scale’

Gerschenkron (1962) advocates that latecomers should take advantage of economies of scale and start with big-scaled production. This runs counter to the idea of the Appropriate Technology (AT) concept, which was originated by Schumacher’s (1973) famous book Small is beautiful. The AT concept, widespread in development studies, often embraces (technology) characters such that small-scale, labour-intensive (in contrast to capital-intensive), energy-efficient and locally (by local people) controlled. The typical AT example is the self-contained (off-grid) solar lamps and/or streetlights (powers e.g. fluorescent) that light up African villages and small towns (in e.g. Sirakorola) without being connected to the national grid system.

With mainly a governance approach, Michael Piore and Charles Sabel (1985), in their influential book The second industrial divide, also raised the issue of scale. In their elaboration, the Fordist style of big-scaled, mass (standardized) production, was (is) facing the challenges of flexible specialization, which is often small scale.

d) ‘Manufacturing-led’ versus ‘service-led’ catching up

Which sector can take the leading role in industrial catching up? This is an old as well as a new debate.
It is an old debate if we trace back to the so-called ‘staple theory’ (of economic growth) that is often linked with the name of Harold Innis (1894–1952), a Canadian economic historian. Emerged in the 1930s and 1940s, Innis’ (1927, 1977) staple theory gives a detailed picture of how the Canadian economy has benefited from the development of a few key commodities such as furs, timber, wheat and cod, as well as from the export of those ‘staples’.

Looking closer and in for example (Three) Industry Revolutions literature (Brynjolfsson and McAfee, 2014; Landes, 1969; Mokyr, 1990), the history of industrialization has always been associated with productivity growth in manufacturing sectors. An alternative formulation is that manufacturing is traditionally considered as having the potential to become a driver for growth and industrialization.

It is also a new debate. In light of the recent rise of India and its software service-led industries (e.g. call centres), which differs from the Chinese manufacturing-led model, debate as regard to paradigms and trajectories of industrialization (and the industrial policies related to that) has become popular (cf. Szirmai et al., 2013). Manufacturing, having strengths in ‘reaping economies of scale, engaging in technological progress and learning, profiting from spillovers to other sectors and providing job opportunities for variously skilled levels of labour’ (Naude and Szirmai, 2013, p. 2), has hitherto been pursued as THE path for industrialization and catching up.

Today, even manufacturing goods are increasingly charged by their service provision (often referred to as ‘servitisation’). The rise of the Indian, Israeli and Irish software industries further brings the service industry-based catching-up path into the research agenda. In Freidman’s (2002) famous ‘the world is flat’, the Indian business process outsourcing (BPO) sector is used to illustrate how beneficial a globalized information society can be. The basic message is that while for example a Swedish academic is sleeping, the proofreading of his/her academic paper is already finished in India and ready to be submitted in the morning when he/she wakes up. The same theory applies to for example tax declaring alike back-office support functions.

From the catching-up perspective, it may be understood that the development of service industry (Indian software industry), rather than that of manufacturing sectors only (Chinese case), can also be a viable path to pursue (for catching up). This could be particularly true with digitization and automation—machines increasingly replacing human beings—becoming the trend.
Question: Is manufacturing-led industrialization still THE path for catching up?

This also questions the very foundations of the concept of Structural Transformation (ST) that was built up by famous scholars such as Kuznets, Chenery and Syrquin, Mellor and Timmer. It is true that the history of industrialization has hitherto witnessed the agriculture–manufacturing–service path in many countries’ catching-up cases. Can the manufacturing part be skipped? We probably do not have answers at this stage.

**The analysability**

What does the theory do and what does it not do when using it in analysing modern cases?

**The insufficiencies**

Which catching-up factors, conditions and paths have played a role is essentially an empirical question open for research. The views presented above in Section 2 can only suggest the awareness on what might (not) be in effect.

Moreover, in modern case analyses, the insufficiencies (of the theories) often appear and that is also the very logic why theories get developed over time.

To exemplify, in the trajectory discussion, challenges brought up by climate changes are not discussed, although it is important. The path(s) of industrial catching up may be modified because current engineering frontiers may not exist. The most advanced fossil fuel supported combustion engine technologies—widely used in making airplanes and cars—may not be relevant any more, if climate change is ever to be seriously considered. The rate and direction of catching up are also affected. For example, the Chinese (automobile) manufacturing sector may have to reduce supply to handle problems caused by pollution (which in turn is caused by fossil fuel-based energy consumption).

In a concrete case analysis, it is, therefore, important to keep in mind that many new challenges are potentially to be endogenized, as their impact is not only an issue of (quantitative) outcome but also an issue of trajectory (of industrial catching up).

Moreover, two important industrial trends—missing parameters in the above-mentioned classic arguments—may be included if applicable: modularization and digitization.
Modularization refers to a trend that the product, as well as production process, is capable of being ‘modularized’ (Baldwin and Clark, 2000). This has an impact not least on design industries, in for example semiconductor design (Ernst, 2005; 2015) and additive manufacturing (with 3-D printing).

Digitization is another trend with growing importance to the modern transformation of many industry sectors, due to the generic function of many ICTs.

Technologies under the ‘information technologies’ umbrella are numerous: the Internet of things, cloud computing, 3-D printing. Application cases are even more pervasive: Airbnb, Spotify, driverless car ... The basic message is that ‘machines’—capable of complex communication and pattern recognition (Brynjolfsson and McAfee, 2014)—are increasingly helping as well as replacing human beings.

The concrete implications of the trends just mentioned are essentially empirical, namely to be analysed case by case.

To summarize, there are many new conditions to be met if a phoenix is to rise from the ashes and that are not discussed in the classic literature presented in Section 2. It is the job of the actual researcher to lift up those conditions, to give a finely grained analysis.

**The cases**

Again, which catching-up factors, conditions and paths have played a role is essentially an empirical question. I briefly illustrate this point with two cases:

**Case One: Is Vietnam catching up? If so, where is the technological entry?**

Vietnam, with a population size similar to that of Germany (more than 90 million), is now enjoying the world’s second-fastest growth rate, 7% on average since 1990. While evidently there is macroeconomic growth, is there also an industrial catching up? If so, where—in which technological field—is it entering?

Factors: what are the important factors here?

At the human capital front, Vietnam, a (former) communist state, has always addressed the importance of education. That, in turn, conditions the absorptive capacity needed for industry transformation. Still, has Vietnam managed to ‘commercialize’ these inherited competences? If so, in the software industry as the Indians have done? Or in the manufacturing sector as the Chinese have done?
At the institution front, Vietnam has had open-trade policies since the 1990s and attracted many FDIs. Trade now accounts for roughly 150% of GDP, according to The Economist (Aug 6th, 2016) indicating a high level of openness. High-tech parks and open zones are additional innovation institutions created for that purpose, namely absorbing foreign knowledge. Still, to what extent has advanced technological knowledge been generated under this facilitation? Examples? Cases?

Paths and conditions: digitization, mostly automation throughout manufacturing sectors, is argued to have wiped out the possibilities for Vietnam entering into labour-intensive manufacturing sectors (e.g. clothing) like China once did.

We see therefore potentials and windows of opportunities for Vietnam to catch-up. If so, which sector would take the lead in this catching-up process is however still an empirical puzzle to be further solved.

Case Two: Is Wechat (a social medium) catching up (and leapfrogging) Facebook in service provision?

WeChat–WeiXin in Chinese—is a mobile terminal-based social media platform. It is widely used in the Chinese community, both in China as well as by overseas Chinese. It had over 700 million active users by 2015. In a departure from its social media origin, WeChat now is widely used by Chinese consumers in ordering and paying (Didi) taxi services, booking restaurants, buying train tickets, making hospital appointments and paying electricity bills. WeChat QR (quick response) codes are replacing business cards; WeChat is replacing plastic bank cards in payment. Email has never functioned as an official way to communicate, but WeChat does now, in many businesses in China.

Is WeChat a case of leapfrogging (to e.g. Facebook) in social media service provision? If so, where (in which field), how and why?

Factors

Here the network externality mechanism typically applies. The consumer base, rather than other forms of capital, is the biggest asset for the company Tencent (the developer of WeChat). This is a factor of competitiveness that Facebook also has. Facebook’s customer base is perhaps even more international.

However, currently, Tencent has managed to form a huge consortium, collecting many (Chinese) services providers around this application. Didi taxi—the Chinese version of Uber—is one of them. Many hospitals and their mobile-based appointment systems are also included. The co-operation
demonstrates a symbiosis between WeChat and other forms of Internet-based service providers. This kind of collaboration—services beyond service media—is a factor of competitiveness that Facebook perhaps does not have.

At the institution front, many social habits—‘informal institution’ in North’s (1990) words—are taken into consideration in this new service provision. This is illustrated by the ‘red envelope’ app, which, in 2014, transferred 32 million packets of digital cash during the New Year celebration (The Economist, 6th Aug 2016).

**Conditions and paths**

China is a market in which smartphones outweigh PCs in Internet surfing and communications. PCs have never managed to become the dominant communication terminal, unlike the case in many OECD countries. This is a trait that is shared by many developing countries, namely mobile technologies dominate Internet access. Moreover, there is a full-blown network externalities effect here, supported by the scale of subscriptions.

This is not a high-tech sector, perhaps, but WeChat shows a pervasiveness of digitalization also in China. WeChat’s former PC-based form—QQ—may be considered as an imitation of Facebook. But its mobile-based form—WeChat—certainly is not.

Still, there are many unknown factors. Although WeChat seems to provide more services than other social media platforms do, it is hard to say, at this stage, whether it is leading the service provision (leapfrogging Facebook) or not. It is also an issue of measurement. We may need to break down to sub-service sectors to get a more detailed picture. While Facebook may have advanced the social media services at many dimensions, WeChat, as a latecomer, may have advanced other niches.

**Managerial and policy implications**

Some of the implications have already been discussed in the above subsections.

**Managerial implications**

The managerial approach to industrial catching up has been centred on strategies, (technological) capabilities and innovations at the firm level. This chapter provides a wider perspective, mainly through introducing studies conducted also at the macro (country) level and at the industrial (meso) level, whereas a firm-centred approach often treats them as contexts.
As I see it, despite the importance of (innovation) strategies, a firm’s catching-up potential—alternatively that of forging ahead and/or falling behind—cannot be assessed with detachment from its national/regional environment, and/or from its sectoral-specific opportunities. Some constraints, such as those related to the institutions, are country-specific and/or regionally confined. That, in turn, affects firms’ internationalization strategies (on trade and innovation).

It is important for managers to understand what is a firm-specific constraint (hindering forging ahead) and what are sectoral-specific (technological) opportunities that are also open to competitors. The structure of this chapter, by sorting out the arguments on factors, conditions and trajectories, can help managers at this front.

Policy implications

Industrial catching-up studies—embracing subjects such as choice of technologies (cf. Sen, 1960), skills and capabilities, entry barriers and trajectories of technology development—are inherently policy issues too. These are issues often discussed in for example ‘dependency theory’ (Gunder, 1972) with regard to the causes of underdevelopment, for example.

Suggested further readings

Two strands of the literature may be suggested as further readings: a) the low carbon trajectory on catching up (e.g. Niklas Stern, 2015); and b) the upgrading indicators that are essentially connected with for example the Global Value Chain (GVC) (Gereffi, 1999; 2005) literature.

On methodological considerations and the use of sources in this chapter

This chapter, like many other chapters in this book, is mainly a literature review chapter. There are interpretations, as well as perspectives, from the author, on the existing literature. So the first category of sources is the literature, the theories.

The majority of the theories, however, were generated after the emergence of a group of interlinked empirical phenomena. The theories are therefore often a laggard and not sufficient in explaining (some of) the causations. The author’s own research, including the on-going studies, as well as the author’s reflections and interpretations are therefore complemented as the second category of sources for this book chapter.
The sources used here, therefore, have a mixed character, namely including first-hand field data, second-hand reports/studies and the author’s own interpretations. Arguably, all of the review-like chapters share this character.

No source of data is problem-free. In development studies, the reliability (also validity) of the statistical data collected from developing countries is often questioned. Moreover, terms have very different meanings. For example the term ‘unemployment’ connotes very different things in the context of developing countries versus industrialized countries (Myrdal, 1967). Quantitative data collected there, including the survey data, would encounter great challenges at this definitional front.

I consider that it is important to go beyond the ‘symptom’ and to reach the ‘diagnosis’. A famous quantitative economic historian, Angus Maddison (et al., 2008) and his colleagues, for example have explored how different accounting systems have evolved over time with a bearing on the statistical terms. As I see it, it is also important to see how that evolution affects our usage of the statistical data.

References


116-131.

Arora, A.; Gambardella, A (eds). (2005) From Underdogs to Tigers : The 
Rise and Growth of the Software Industry in Brazil, China, India, Ireland, 
and Israel. Oxford University Press

Baumol, W. J., Blackman, S. A. B. and Wolff, E. N.(1989) Productivity and 
American leadership. Cambridge, Ma: MIT Press

Bell, M; Pavitt, K (1995) The development of technological capabilities. in 
Trade, Technology and International Competitiveness, EDI (Economical 
Development Institute of the World Bank) Development Studies. 69-101

Berg Publishers Limited.


Cohen, W., Levinthal, D. 1990. Absorptive capacity: a new perspective on 
learning and innovation, Administrative Science Quarterly, 35 (1) 12 
8-152

Corden, W,& Neary, J. (1982). Booming Sector and De-industrialisation in 

Oxford University Press.


Fagerberg, J.; Mowery, D; Nelson, R. (Eds.) The Oxford Handbook of 

Fagerberg, J. & Srholec, M. (2009). Innovation systems, technology and 
development: unpacking the relationship(s). In Lundyall, B; Joseph, K.J.; 
Chaminade, C.; Vang, J. (Eds). Handbook of innovation systems and 
developing countries. Cheltenham: Edward Elgar. 83-115

Freeman, C. (1994) “Technological Revolutions and Catching-up: ICT and 
the NICs,” in J. Fagerberg, N. von Tunzelman, and B. Verspagen (eds), 
The Dynamics of Technology, Trade and Growth. London: Edward Elgar.


Gore, C. (2000), the rise and fall of the Washington Consensus as a paradigm for developing countries. *World Development* Vol. 28, No. 5, pp. 789-804,


Rodrik, D. & Subramanian, A. (2003), The Primacy of Institutions (and what this does and does not mean), Finance & Development, June 2003


Sach, J. (2003), Institutions Matter, but Not for Everything: The role of geography and resource endowments in development shouldn’t be underestimated, Finance & Development, June 2003, 38-41


Veblen, T. (1915), Imperial Germany and the industrial revolution, New York: Macmillan.


12: Policy Instruments for Energy Efficiency
Rurik Holmberg

In this chapter I will discuss knowledge as a key component for energy efficiency in industry. The role of energy efficiency has recently been highlighted as a means for achieving such goals as reduced emissions of green-house gases or to maintain competitiveness. I will argue that the key for successful implementation of energy efficiency is knowledge as distinct from plain information, because it is something that has to be created in-site. There exists no “one size fits all”, but instead “learning by doing”.

In this chapter, based on experiences from the Swedish programme for energy efficiency in large industries, I will argue that left only to market forces, the full potential for energy efficiency will not be reached. Instead, energy efficiency has to be seen embedded in a broader societal context.

Introduction

In the energy sector the concept of free markets is rather recent. Until the liberalization of the markets in the 1990s, government intervention could be described as “the visible hand”, i.e. direct government-control was the predominant arrangement. Therefore, the recent focus on policy instruments in the energy sector can also be interpreted as the return of the government after a few years of alleged relative absence. This aspect has become increasingly visible when governments are called upon to mitigate harmful side-effects, such as emission of green-house gasses, or to promote energy efficiency. Thus, I will argue that direct government control of the energy sector has been replaced by proxy control through policy instruments. Specific knowledge being the key for achieving goals such as improved energy efficiency or reduction of emissions of green-house gasses, one of the main challenges for policy designers becomes how to promote the creation of knowledge through policy instruments.

By the concept of policy instruments. I refer to some sort of government interference in the functioning of the economy with the aim to influence developments or outcomes in a certain direction without actually controlling
the process. Usually the explicit logic of policy instruments is to remedy some real or perceived failure in the market, i.e. left to itself the market is not expected to produce an optimal outcome.

A well-designed policy instrument will, of course, induce a process to achieve the intended outcome while at the same time being cost-efficient and without significant negative side-effects. Setting a goal is a crucial component for the effectiveness and efficiency of a policy instrument. The former concept refers to the ability of a policy instrument for reaching the goal, while the latter focuses on the cost-efficiency component (Hellmark & Jacobsson 2012).

In addition to this, it needs to be pointed out that what works well in theory will not always be appropriate in practice. Thus policy instruments can turn out to be costly in comparison to the effect achieved, they may have no effect at all or, in worst cases, have an effect contrary to the one intended. A well-functioning policy instrument can, on the other hand, be assumed to give rise to a kind of a multiplier effect by triggering innovative solutions from those targeted which in turn strengthens or increases the pace reaching the original goal. The concept of innovation is beyond the scope of this chapter, but it suffices to say that leaving innovation in the energy sector to the market forces alone does not necessarily produce a different outcome in terms of innovation of new energy-efficiency-related hardware than relying on government-controlled innovation (Markard & Truffer 2006).

Analysing changes in industrial or technological systems with an emphasis on the actions of, or interaction between, government, firms, or some other corresponding entity, is usually referred to as an institutional approach. This could be referred to as “the context or framework of economic activity” (Carlsson 1997). Institutions can also be perceived as “the rules of the game” and can be divided into formal and informal institutions, where the former are codified, e.g. laws, while the latter are for instance cultural norms in society (North 1990). Another way to see institutions is a guarantor of predictability in society (Nooteboom 2002), which is supported by the observation that institutions are less prone to change than technology (Matthews 1986). However, an interesting question thus becomes whether technology develops by and large independently from institutions, or can institutions induce technological change? In this article, I will discuss to what extent we can observe the core element of technology, i.e. knowledge, to be directly influenced by institutional circumstances.

**Aims**

In this chapter I will shed light on the actual outcome of a policy instrument for energy efficiency in industry, especially in contrast to the ex-ante
justification of the instrument, thus adding new perspectives to the discussion on how government intervention in the energy sector can be designed today.

In the energy sector, policy instruments inducing measures in either the supply or the demand side can produce similar outcomes partly because of their oftentimes immediate interdependence (e.g. electricity is produced at the same time it is consumed) and partly because energy as such lacks substitutes, i.e. the source of energy can be changed, but not energy itself. Therefore, targeting only one or the other side can produce equally desired outcomes as targeting both sides simultaneously. Hence, in this paper I will discuss both aspects of policy instruments interchangeably.

The effect of policy instruments is a relatively new field of study and therefore methodological tools have been developed only in recent years (Harmelink et al. 2007). In addition to this, there has been a tendency to focus on measurable outcomes of environment-related policy instruments (e.g. number of new patents or changes in abatement costs), while less attention has been devoted to “soft” effects such as organisational changes or attitudes (Vollebergh 2007; Palm & Thollander 2010).

By focusing on a single case, the Swedish programme for energy efficiency in energy-intensive industry (PFE), I will argue that the actual outcome of a policy instrument may differ significantly from what was originally intended. I will claim that a policy instrument can set processes in motion which enhance energy efficiency, but which were not planned for or even counted with, such as organisational or attitude-related impacts, but perhaps more crucially through the generation of site-specific knowledge. Thus, the purpose of this chapter becomes to contribute to the discussion on unintended, but actual impacts of policy instruments.

I will begin this chapter with a discussion on the role of the government in the energy sector, followed by a discussion of the role of knowledge in industry and a discussion of the theoretical foundations for policy instruments. These aspects are then interwoven in an analysis of the Swedish policy instrument named programme for energy efficiency in energy intensive industry, PFE. Focus will be on the electricity sector, but because in this case it cannot unequivocally be separated from the energy sector in general, some interchangeable use of the terms will occur. 65

---

65 For example, PFE was originally intended to cover electricity only, but developments quickly made a strict separation obsolete, i.e. all energy efficiency was included (Energimyndigheten 2015c)
The changing role of the government

The government has been a key player behind the development of both industries and markets throughout the modern world (North 2005; Helpman 2004; Lindblom 2001). This holds particularly true for general purpose technologies (i.e. technologies which can be used for several purposes), thereby defining the direction of their development, such as electricity generation (Hughes 1983; Helpman 1986).

Magnusson (2005) argues that the government’s tool-box has historically consisted of general regulations, investments in infrastructure, macro-economic policy, stimulation of demand, providing venture capital, and guaranteeing the continued operations of enterprises in times of economic crisis. Thus, according to Magnusson, the assumption that the market would have been the foremost driver behind industrialization can be questioned on good grounds.

Energy, and in particular electricity generation, has traditionally been perceived as an area where government intervention is welcome or even required. A century ago, the development of power stations and the power grid in Sweden took place under government auspices, if not direct control (Fridlund 1998). The electricity sector was dominated by the government-controlled company Vattenfall, while the rest consisted mainly of municipal companies (Damsgaard & Green 2005). In 1996, market mechanisms were introduced on a broad scale in the Swedish electricity sector because electricity had come to be regarded as almost any other commodity (Högselius & Kaijser 2007). However, recently the role of the government has been highlighted anew as a consequence of actions for climate change mitigation and the ensuing new policies for renewable energy and energy efficiency. This is reflected in the role assigned to national governments in for example EU Directive 2009/28/EC (“the renewables directive”) and Directive 2012/27/EU (“the energy efficiency directive”), which are obliged to set certain targets and to carry out a number of measures in order to promote goals in these areas. In particular, Directive 2012/27/EU even requires specific action, or measures, from the member states’ governments. 66

Furthermore, international climate change policies require coordination, where the national governments are the obvious nodes

In Sweden, the character of government intervention in the energy sector is significantly different today from a century ago, but ideas of an energy market free from government intervention have turned out to be at least partially illusory, especially regarding electricity. Against this background, the crucial issue becomes how to identify those policy instruments that produce the desired outcomes (e.g. reduction of emissions of green-house gasses and

66 Article 7 of Directive 2012/27/EU
improved energy efficiency). Nevertheless, while acknowledging that reaching such outcomes requires new knowledge, the arguments for intervention are usually justified exclusively with a reference to the operation of the markets forces alone and thereby not to dynamic effects such as knowledge creation (Energimyndigheten 2013).

**The role of knowledge**

Industrial sites differ from each other in many important respects. Therefore, one prerequisite for energy efficiency in industry is knowledge of the site-specific circumstances and, based on this, in-depth understanding of which measures can generate the most cost-efficient results. In particular, the presence of knowledgeable staff and a long-term strategy appears to be conclusive for implementing successful measures for energy efficiency (Thollander & Ottosson 2008). The aggregate knowledge in a company exceeds the sum of competencies and capabilities of staff and equipment and knowledge is maintained by using it, in other words there is a constant process of “learning by doing”. Knowledge is, as a consequence, constantly changing, which is caused by a number of factors, including government-ordered R&D. Because knowledge resides within the firm as an entity, it is manifested through intra-firm routines. There exists no “book of blueprints”, no “chief engineer” possessing all relevant knowledge, but only firm-specific ways of carrying out tasks (Nelson & Winter 1982).

The knowledge required for energy efficiency measures in industry is, to a large extent, not codified, but has to be constantly re-created (Palm & Thollander 2010), and can therefore be perceived as “tacit knowledge”, a concept defined by Michael Polanyi (1958/78) as “knowledge of the particular circumstances of time and place”. Those who have developed this sort of tacit knowledge can be assumed to have such an understanding of the specific industrial site so as to take appropriate action when there is a change in circumstances. This knowledge can therefore not be treated as information and it does not lend itself to measuring. Measurable indicators, such as investment in R&D can only be approximations (Foray 2004). According to Eliasson (1990) organizations, including firms, are the actual engines of knowledge-generating in society. This view is shared by von Tunzelmann (1997) who perceives firms as knowledge-processing units.

Nakicenovic (2002) ascribes knowledge obtained through “learning by doing” even greater importance than external innovations in the case of reducing emissions of carbon dioxide. This observation appears to be relevant when it comes to energy efficiency, too (Broberg Viklund 2015). Albeit new technology plays an important role in enhanced intra-firm capacity for energy efficiency, the key point appears to be the development of new routines (Energimyndigheten 2015c).
Remedies for market failures

The concept of market failures has been one of the principal formal arguments for introducing policy instruments in the Swedish energy sector, according to the Swedish Energy Agency (Energimyndigheten 2013)

According to neoclassical economic theory a so called perfect market can be envisioned, which produces the optimal outcome in terms of cost and utility. This construction, which few economists would claim actually exists, can be used as a sort of a reference for what could theoretically be possible to achieve, or an ideal state. Therefore, empirically observed divergences are characterized as market failures, i.e. for some reason the market is not yielding optimal results, e.g. profitable measures for energy efficiency are not being implemented (Radetzki 2005). The assumed reasons for market failures can be lack of information, negative externalities (i.e. unintended and unwanted side effects) or asymmetric information, by which is referred to situations where one part has more insight than others and is able to use this to his or her advantage.

Market failures can, according to theory, be remedied by policy instruments. Thus, for instance, lack of information can be remedied by information campaigns. In this view, one policy instrument should target one specific market failure, because otherwise the instruments would overlap one another and thereby reduce the overall cost effectiveness. In many cases it is more or less obvious which instrument suits a certain market failure, as is for example a tax on emissions (an externality) if the intention is to curb emissions.

However, in reality this simple approach is not always applicable and often the market failures addressed tend to be both complex and interwoven. Therefore, more sophisticated approaches are needed in practice. In addition, the sequence in which policy instruments are introduced may be decisive for the outcome (Bye & Bruvoll 2008).

Classification of policy instruments in Sweden

In Sweden, policy instruments have usually been divided into administrative, informative, economic, and research and development (R&D). Administrative policy instruments are for example laws and regulations, while taxes are typical economic policy instruments. An informative policy instrument is the network of municipal energy and climate advisors in Sweden.

This classification has recently been perceived insufficient, not the least because it does not take into consideration processes and activities carried out by those targeted by the instruments. Therefore, it has become increasingly common to add engaging instruments and physical planning as separate categories. Networks for industry branches or technology procurement groups
are examples of engaging policy instruments while municipal planning of transports illustrate the concept of physical planning (Energimyndigheten 2013).

**Some Swedish policy instruments**

In Sweden there are currently several policy instruments in place in the energy field. Illustrative examples are the electricity certificate system, the aim of which is to enhance renewable electricity generation, taxation of vehicles based on emissions of carbon dioxide, and technology procurement groups such as BELOK, where several owners of commercial premises collectively purchase new technology whereby a downward pressure is put on prices, thus making new energy efficient technology more affordable. Another policy instrument is the building regulations by the Swedish National Board of Housing, Building and Planning (Boverket). On the EU level there are for instance regulations on energy labelling of certain products (Directive 2010/30/EU) or the directive on energy performance of buildings (EPBD, Directive 2010/31/EU), which have been transposed into Swedish law.

**Policy instruments for industry**

One example of what appears to be a typical market failure is energy efficiency in Swedish industry. The potential for cost efficient measures is estimated to be significant, but still only a minor part of this potential is realized. The consultancy SWECO (2014) estimates in a report to the Swedish government that this potential will reach almost 40 TWh by 2030, from which only approximately 12-13 TWh is realized spontaneously, i.e. without specific policy instruments or other external actions. Finding the reason for and addressing the under-utilization of this vast potential for cost efficient measures for energy efficiency has naturally become of great interest for policy makers as well as managers of industrial sites.

Nevertheless, there are several policy instruments in place targeting industry. One such policy instrument is the European carbon emission trading system (EU ETS). In Sweden there is also a number of voluntary networks in which industry in some branches co-operate to develop practices for energy efficiency measures. These networks have proven rather successful in disseminating information, not the least because companies tend to consider information stemming from other companies in the network as particularly trustworthy (Rohdin, Thollander & Solding 2006). It remains to be seen whether knowledge can be transferred through networks, too.

**Engaging instruments for knowledge creation**

As I have discussed earlier, there exists knowledge which cannot be disseminated through information campaigns directed at industry, and
consequently no central actor, such as a government agency, can reach the targeted companies in an appropriate manner by simply informing. Market failures often incorrectly identified as information failures will not be properly addressed unless the underlying cause is adequately identified. Thus, the lack of action for addressing the vast pool of low-cost measures for energy efficiency in industrial companies due to insufficient knowledge is not caused by information failure, but rather that appropriate knowledge has not emerged or been created. In recent years, this question has increasingly been identified as a shortcoming of available policy instruments (Palm & Thollander 2010). The emergence of so-called engaging policy instruments in the last few years should be seen against this backdrop.

Recently, there has been growing interest for a group of complex policy instruments often referred to as white certificates or energy savings obligations. This group of policy instruments is based on the assumption that energy companies (either energy sales companies or network operators) should carry some responsibility for making energy efficiency measures to be carried out among their customers, i.e. the final consumers of energy, as for example stipulated by article 7 of Directive 2012/27/EU (Energimyndigheten 2015b).

This type of policy instruments targets several market failures while at the same time they can be combined with other policy instruments to obtain their full potential effect. The underlying assumption is that energy companies will develop their business models towards offering their customers services instead of maximizing the amount of energy delivered. This reasoning is sometimes simplified to the statement that customers don’t actually want a kWh, but the services energy can provide. Simply put, the less energy is required, the better. There is an underlying idea that white certificates would thus enhance knowledge creation. (Energimyndigheten 2015b).

Currently, there are no concrete plans to introduce white certificates in Sweden. In countries where this policy instrument has been in place for some period of time, there is however, relatively little evidence that the business models of energy companies would have altered significantly. On the other hand, it is possible that these policy instruments have not taken full effect yet, because they have been introduced relatively recently. There is some evidence from Denmark that changes in the pattern of energy use among end-users can be observed (Energimyndigheten 2015b). If this turns out to be the case, white certificates would be a policy instrument targeting the supply side but producing an outcome on the demand side.

The key to understanding the functioning of white certificates is that site-specific knowledge of energy efficiency is developed in co-operation between energy companies (who possess in-depth knowledge of energy efficiency as such), and the end-users, (who possess in-depth knowledge of their facilities).
In the early 1990s, Sweden introduced a tax on the emission of carbon dioxide. This tax is for practical reasons often lumped together with the energy tax, which was introduced in the 1950s as a purely fiscal tax. This policy instrument is usually perceived as the backbone of Swedish energy efficiency policy, a claim which is supported by the fact that the Swedish government has reported only the effects of these taxes to the European Commission in connection with requirements following from the energy efficiency directive (Regeringskansliet 2013). However, companies participating in the European Emission Trading System (EU ETS) are exempted from the carbon dioxide tax (but not from the energy tax). This means that most of the emission-intensive industries are not covered by the tax and, as a result, an analysis of the dynamic effects of the carbon dioxide tax would be clearly limited in scope.

Taxes on energy and emissions of carbon dioxide stimulate conversion from fossil to renewable energy sources, but also reduction of overall energy use. Each end-user is left with the decision to what extent this is made and by what means. Knowledge of available options to achieve reductions can therefore be said to be rewarded through lower taxes and lower energy bills.

Taxes are often favoured for being technology neutral, i.e. no authority dictates exactly how to proceed with the task of reducing emissions or energy use. On the other hand, the introduction of taxes does not necessarily lead to any of these goals, because end-users may prefer to pay the tax instead of carrying out any measures. In that case, it is obvious that taxes would not have any significant impact on knowledge creation, but instead become only an additional cost.

In brief, comparing white certificates and taxes, it could be claimed that taxes have the advantage of being technology neutral, but “passive” in the sense that they do not particularly promote knowledge-generation. White certificates tend, on the contrary, to stimulate the search for site-specific solutions, i.e. knowledge-generation, but they also contain a risk that focus is laid on “low-hanging fruits”, i.e. simple and rather obvious measures, unless the target for the entire white certificate scheme is sufficiently ambitious (Energimyndigheten 2015b).

The case of PFE

EU’s energy tax directive from 2003 (Directive 2003/96/EC) stipulates a minimum tax on electricity, approximately 0.005 SEK/kWh. This tax was regarded as problematic by both the Swedish government as well as the Swedish energy-intensive industry, because of the challenge it posed to the competitiveness for some of Sweden’s crucial industry branches, such as pulp and paper or steel. As a countermeasure, the government introduced the programme for energy efficiency in energy-intensive industry (PFE), which was launched in 2004. PFE, which is based on voluntary participation,
participating companies a waiver on the electricity tax in exchange for active measures to improve energy efficiency, including the introduction of an energy management system (EMS) provided there was none in place before (Energimyndigheten 2015a).

It should be emphasized that although the focal question for this policy instrument was electricity, it was soon interpreted to encompass all energy consumption. Thollander & Ottosson (2010) arrive at the conclusion that prior to PFE, less than half of Swedish energy-intensive firms had a systematic approach to energy efficiency by the way of energy management. This might be a consequence of energy efficiency not being a part of their core business and thus paid less attention to. However, such a finding would on the other hand partially support the estimates that there exists a significant energy efficiency gap.

Nevertheless, PFE has been gradually wound down starting in 2014 after the EU Commission judged it being in breach with EU’s rules on state subsidies to industry. No new companies can participate any longer, but those who joined at a late stage may continue implementing the programme until end-2017.

Participating companies have reported their actions for energy efficiency to the Swedish Energy Agency, and consequently there exists a comprehensive data material for analyses of the effects of PFE. A number of evaluations have been carried out, which are far from unanimous in their conclusions.

The Swedish National Audit Office (Riksrevisionen 2013) concludes that the results of energy efficiency measures following PFE remain unclear and that they may equally well stem from other circumstances, such as higher prices on electricity in general. The National Audit Office compares electricity consumption in companies participating in the PFE with corresponding figures for non-participating companies and draws the conclusion that there is no difference in neither absolute terms (total electricity consumption) nor in relative terms (electricity use related to value of production). Nevertheless, the National Audit Office finds that the value of investment in energy efficiency made by participating companies broadly corresponds with the total amount of tax waivers for these companies during the first five years of PFE (until 2010). Total tax waivers amounted to 750 million SEK, while investments in energy efficiency are estimated at 708 million SEK. There is only little evidence of new technologies for energy efficiency emerging in these industries, thus excluding the explanation that the investment in question would have resulted in increased R&D.

Following from this, the National Audit Office notes something rather paradoxical. Investments for 708 million SEK have apparently not produced any measurable result. The National Audit Office notes that the participating
companies have, in addition to direct measures for energy efficiency, also introduced Energy Management Systems (EMS) in accordance with the requirements in PFE, but does not elaborate further on this issue.

Stenqvist & Nilsson (2012) assess the magnitude of energy efficiency measures stemming from PFE at an annual 1.45 TWh, while the net effect is estimated to be between 0.69 and 1.01 TWh annually (i.e. discounting the effects the electricity tax would have had). In addition to this, measures targeting the use of fuels and heat can be observed to amount to an annual 1 TWh. This additional effect is ascribed by Stenqvist & Nilsson largely to the introduction of EMS. So far, the participating companies have focused on “low hanging fruits”, i.e. measures that provide a pay-back time of 1.5 years on average. This means that there probably still remain profitable measures to be found, at least measures with longer pay-back times.

In a study based on interviews, Broberg Viklund (2015) reveals that among participant companies of PFE, the identification and localization of heat flows tend to be split between several parts of the companies with nobody actually in charge. Therefore, appointing an energy manager addresses this aspect of energy efficiency. However, a bottleneck appears when new investment to remedy efficiency shortcomings is required. Therefore, Broberg Viklund (2015) concludes, a specific policy instrument is needed in order to stimulate an efficient management of heat flows.

An EMS, in principle the current international standard ISO 50 001:2011 is described by the Swedish Standards Institute, SIS (the Swedish counterpart in international standardization work) as “a system aimed at supporting an organization to implement a systematic approach to improving its energy performance, including energy efficiency, energy use and energy consumption.” Moreover, the standard “specifies demands to be implemented on energy use and energy consumption, including measuring, documentation and reporting, construction and design as well as routines for procurement of equipment, systems, processes and staff contributing to energy performance”.

ISO 50 001 is the latest standard, but at the time PFE was launched in 2004 there existed no international standard for EMS, but instead environmental management systems were used. Today, an EMS is in place in basically all of the companies concerned. Participating companies commonly perceive that the implementation of an EMS has changed the ways issues related to energy efficiency are managed within the companies. In particular, it puts a designated person in charge and made energy efficiency a topic in its own right at board meetings, thereby legitimizing actions for energy efficiency (Thollaner & Ottosson 2010; Broberg Viklund 2015; Energimyndigheten 2015a).
Thus, it can be concluded that fulfilling the requirements mentioned above (as the Swedish National Audit Office has found), the participating companies must have carried out significant activities promoting energy efficiency, but this is not visible in available statistics on energy use. Therefore, the question is what has actually taken place. A speculative answer would be that these companies have avoided an increase of energy consumption. It would, nonetheless, be difficult to prove such an interpretation.

It could also be possible to challenge the method used by the National Audit Office by highlighting the fact that there exists no adequate reference group to the companies participating in PFE, because basically all companies that could be defined as energy-intensive actually participated in the PFE at some point. Those companies that did not participate had simply much lower energy consumption from the outset and are therefore not comparable.

Yet another explanation could be that participating companies gradually diverted to some extent from the original idea of the programme by carrying out efficiency measures targeting not only electricity but all energy consumption (Stenqvist & Nilsson 2012). This assumption seems to be supported by the fact that the participating companies showed a remarkably stable relation between energy use and production value during the sharp downturn 2008-09. The participating companies are usually operating processes in which total energy use cannot be rapidly altered to any significant extent. Therefore, a reduction in output with the actual production process remaining unchanged would indicate that energy efficiency measures might have produced this outcome. However, these findings need further support before far-reaching conclusions are drawn. (Energimyndigheten 2015a).

It might, however, be difficult to find data supporting any of these tentative explanations. At least partially, the lack of hard evidence for energy efficiency progress could naturally be explained by so-called rebound effects, by which is meant that gains through energy efficiency measures lead to increased use in some other part of the system. Estimating the size of the rebound effect has proven highly complicated, not the least because it tends to vary on a case-by-case basis, but according to Broberg (2011) it can be of significant size in particular in energy-intensive industry.

**Managerial and policy implications**

Interpreting the existence of an energy efficiency gap in industry as a consequence of market failures, as is the prevalent view among Swedish policy makers, has its merits, but this interpretation is obviously not sufficient for addressing the underlying problems properly. Especially when limiting market failures and the corresponding correcting policy instruments to the
traditional categories (administrative, economic, informative and R&D), progress in energy efficiency in industry will remain unsatisfactory.

Knowledge of site-specific circumstances is the key to energy efficiency in industry. By knowledge is referred to something more comprehensive and complex than plain information, and therefore a policy instrument targeting knowledge as a way to enhance energy efficiency appears to have greater likelihood for success than others. This is because the knowledge needed does not exist yet, but has to be created or developed. However, this connection seems not to have been fully understood until recently.

Proper addressing of energy efficiency in industry thus requires approaches beyond a simplified model in which market actors react to external stimuli. Instead, evidence suggests that awareness of the actual potential for energy efficiency might be limited and the tools to achieve it underdeveloped. Therefore, specific knowledge needs to be created in-site (compare with Polany’s “knowledge of the particular circumstances of time and place”). In order to set the knowledge-generating process in motion, outside intervention by means of policy instruments appears to be required, but the design of the policy instruments needs to encompass also the active involvement of staff, i.e. it is not enough to levy fees or to produce information. Knowledge being one of the crucial components of any firm, i.e. it resides within the firm as a whole and it is manifested in the routines (Nelson & Winter 1982), implies that a policy instrument inducing a change in the routines could be a key to accessing the full potential for energy efficiency in industry.

Studies indicate that the policy instrument PFE for energy intensive industry has resulted in significant financial resources being directed to energy efficiency purposes in the participating firms. However, corresponding measurable results are still wanting, which has raised several questions.

Through the introduction of EMS, energy-intensive companies have implemented organizational measures needed in order to make energy efficiency an integrated part of their normal routines. Energimyndigheten (2015b) and Broberg Viklund (2015) conclude that based on data submitted by the participating companies in the PFE, routines within the companies have actually been affected, not the least because the tax waiver paired with the introduction EMS have resulted in a significantly enhanced status for energy efficiency issues within the companies. The participating companies have reported that as a consequence of PFE, energy efficiency has become an issue for senior management instead of solely an issue for energy managers with an unclear mandate.

However, only introducing EMS in a limited number of companies can hardly explain the investment in excess of SEK 700 million. Excluding previously mentioned measurable actions, one of the remaining explanations is that
investment has been made into site-specific knowledge-generation. This would also at least partially explain the difficulty many studies have in pinpointing what exactly has taken place.

If this observation gains more empirical support, it can be assumed that the effects will only gradually become measurable over the next few years and in that case the PFE could be characterized as an engaging policy instrument although initially designed primarily as an economic policy instrument. From today’s viewpoint, several observations thus point in the direction that the PFE promoted significant investment (a large part of the earlier mentioned SEK 708 million) in what reasonably could be identified as knowledge-generation.

If it turns out to be the case that a certain design of an economic policy instrument actually promotes knowledge-generation and routines within companies, further studies of this subject might generate highly relevant results for future policy making. Nevertheless, it seems clear that significant progress in energy efficiency cannot solely be left to market forces and some corrections through superficial policy instruments. Instead, the most suitable policy instruments should be perceived being an integrated part of a complex web of knowledge creation, where the market forces are seen as only one, albeit crucial, component.

**Suggested further readings**

- The role of institutions in the development of technology is nowadays a broad field. General works on institutional development worth particular mentioning are Nathan Rosenberg’s and L.E. Birdzell’s “How the West Grew Rich” and Douglass North’s “Institutions, Institutional Change and Economic Performance”.
- Knowledge as a key component in the economic development is described in Elhanan Helpman “The Mystery of Economic Growth”
**On methodological considerations and the use of sources in this chapter**

The empirical data in this chapter is to a significant extent based on reports from the Swedish Energy Agency (Energimyndigheten) and other governmental agencies. It should therefore be emphasized that the presentation of data in such sources is never completely independent from the political process.

**References**


European Communities *Directive 2003/96/EC* (Energy tax)
European Communities Directive 2009/28/EC (Renewables)
European Union Directive 2010/30/EU (Labelling)
European Union Directive 2010/31/EU (EPBD)
European Union Directive 2012/27/EU (Energy efficiency)
Den konstuerade världen. Tekniska system i historiskt perspektiv, Brutus Östlings bokförlag
Lindblom, Ch. (2001). The Market System, What It Is, How It Works, and What To Make of It, Yale University


Regeringskansliet N2013/5035/E [Swedish Government report]

Riksdagen (2013) *Energieffektivisering inom industrin – effekter av statens insatser*, RIR 2013:8


Introduction

In this article the development of the Swedish industry, and the industrial dynamics in Sweden, is linked to political and societal changes. One of my ambitions is to demonstrate how the industrialization and the development of a competitive industry in Sweden was formed and has been maintained by a complex interplay between government and business. An interplay between political movements in society, governmental decisions and regulations on the
one hand and business and industrial development on the other. The rules and regulations, as well as the interactions and cooperation between different actors, has changed considerable over time. Shorter periods of great turmoil have been followed by longer periods of a relatively stable growth. The development during one phase has of course also built the foundation and the starting point for the next. In this article I have divided the development 1850-2015 into three different growth periods, each initiated by a period of political and industrial turmoil and great changes in society (This whole chapter, and especially the beginning, draws on 37 articles published in Giertz (Ed) 2008).

The first phase was initiated by a number of liberal reforms during two decades in the mid-1800s. Freedom of business was introduced, an act introducing the principle of limited company liability was passed and the guild system was abolished. Sweden also introduced free trade, a new banking law and the right for any citizen to establish firms. In parallel a political new order brought new men with liberal ideas to the representative bodies. Government initiated investments in new infrastructure, which were combined with local, private investments. The liberal reforms and the new infrastructure set up the foundations for a belated but rapid industrialization in Sweden. Most new firms completely focused on the domestic market but some also became successful export companies in the wood-, pulp and paper-, steel- and engineering industries. During the second half of the 19th century a new entrepreneurial capitalism flourished, which hardly resembled the traditional and stable social structure in the beginning of the 19th century.

The second phase was initiated by the turmoil during the decade following World War I. Full democracy was introduced in Sweden in 1921. But the war was also followed by a deep depression and a tough deflationary monetary policy. Between 1920 and 1922 liquidations and bankruptcies were very common in Sweden. Many company founders and their families lost their businesses, which were taken over by a few financial spheres, which became even fewer after a new depression a decade later. The consolidation of the export industry was accompanied by a massive consolidation of the infra service industries when government increased its grip on the infrastructure, through the acquisition of all local phone associations and telephone companies, the nationalization of all private railroads and the establishment of Vattenfall for regional power generation and later the deployment of a nationwide transmission grid.

It was not only the control over the Swedish industry and infrastructure that was centralized in the 1930s. After the 1936 election the political power in Sweden was gathered in the reformist labour movement’s hands for many decades to come. Leading Social Democrats replaced the socialist dogmatism for a pragmatic cooperation with industry and built close personal contacts with the most important owners and CEOs in Swedish industry. They saw a competitive export industry as the foundation of a future welfare society.
During the 1950s and 1960s Swedish industry harvested major export successes. Sweden's share of world trade rose almost dramatically; Sweden took the lead in terms of economic growth and a welfare society emerged through the close cooperation between central representatives of government, industry and trade unions in a very centrally but informally managed truly corporative society.

In the late 1960s the consensus and close cooperation between government and private industry was put to an end. A young generation of leading Social Democrats, that formed a new government in 1969, took a distance from the former close cooperation with big business. In the beginning of the 1970s thoughts of socialization, governmental enterprises, wage-earner funds, new labour laws and socialistic national economic planning were on top of the political agenda among Social Democrats in Sweden. Thus, towards the end of the second phase the close and informal cooperation between government and industry was put to an end and replaced by more confrontation, legislation and formal bureaucracy. However, the belief in the strong state with a powerful government responsible for developing, controlling and distributing welfare to all citizens in the country rather increased than decreased in the beginning of the 1970s.

The third phase was initiated in the mid-1970s, when the development took a fast new turn with industrial crises and heavily underfunded welfare systems. The crisis was manifested in decreased production, reduced market shares for Swedish industry on the global market, increased unemployment, negative trade balance and low profitability in many industries. The Social Democrats lost the election in 1976 and a new political era, with pending majorities, was born. Focus was once again transferred to economic growth. But this time – not only incumbents – but also new ventures and higher efficiency in the public sector were important targets. Deregulation, which increased the pressure for change in domestic and public activities were carried out.

Since the late 1980s, thus a new pattern emerged – a globalization phase had begun. Trade barriers were taken away and new trade blocs were formed. Meanwhile, the information and communications technology shrunk the world further. Many markets became global. The financial market is global, both manufacturing and service companies are operating on global markets. The companies that provide infra services are nowadays competing on international markets. Questions about corporate locations of activities are constantly under discussion, as well as the exchange of labour in increasingly open economies. Major structural changes permeate the business world. Companies are acquired and merged. Many companies concentrate on their core business and choose to procure services that were previously made in-house. Globalization has also fundamentally changed possibilities for small individual nations, like Sweden, to form national strategies for growth.
Liberal reforms, industrialization and entrepreneurship

Until the mid-1800s Sweden was relatively untouched by the incipient industrialization in other parts of Europe. Liberal ideas that had taken root elsewhere had not left any deep traces in Swedish society, but in the mid-1800s a series of reforms were implemented.

Liberal reforms, a new political order and new communications are introduced

A number of liberal reforms were initiated in 1844 when king Karl XIV Johan was succeeded by his more liberal minded son Oscar I and grandson Karl XV. During two decades a lot of fundamental changes were introduced. Along with a general liberalization Sweden introduced the right for any citizen to establish firms and free trade was introduced; import and export bans were lifted, and both foodstuffs and raw materials were completely duty free. A new banking law was added, which led to the establishment of new commercial banks. The banks attracted capital which was loaned to an expanding industry. Predecessor was A.O. Wallenberg, who founded Stockholms Enskilda Bank already in 1856. The number of commercial banks peaked in 1908. Then there were 84 companies (Lindgren 2008, p. 87). The liberal reform period in the mid-1800s ended with a reform of the political life in 1865. The political new order brought new men to the representative bodies. Conservative groups (nobility, clergy, bourgeois and peasants) lost their former privileges. They were replaced by more liberal forces, in city councils as well as in the new two chamber parliament, which advocated industrial development.

The state also took on new commitments, which were financed by large borrowings abroad. These included development of communications, which improved dramatically from the mid-1800s (Kaijser 2008:1). A novelty was the electric telegraph. A telegraph agency (Telegrafverket) was established in 1853 to build a state-owned national telegraph system. That same year the Stockholm-Uppsala telegraph line was in use for public traffic. Another novelty was the railroad. In 1853 parliament decided to build a nationwide railway. From the start the railway network was planned as a nationwide transport system. Government built a network of trunk lines but left it to private railroad companies to build the ramified connections. In 1880 there was a total of 6,000 kilometers railway in the country, of which 2,000 were owned by the state and controlled by Swedish Rail (SJ). So in constructing the nationwide railway system governmental initiatives were combined with local private entrepreneurship. For telephony, as well as electricity, on the other hand the exploitation started almost invariably by local initiatives.

In 1880 the Stockholm Bell Telephone Company established the first telephone network in Sweden with just over one hundred subscribers (Kaijser 2008:1). The Bell Company sold telephone networks to public institutions and
companies. Bell applied the same concept in many other cities in the US and in Europe. In some countries the Bell Company had a patent, but not in Sweden. In Stockholm the Bell Company was challenged. The engineer and entrepreneur Henrik Tore Cedergren felt that the phone should also be aimed at households. He established a competing telephone company in 1883, Stockholms Allmänna Telefonaktiebolag (the Stockholm Public Telephone Company), which purchased technical input from a newly founded engineering company, Telefonaktiebolaget LM Ericsson, which also supplied some spare parts to the telegraph agency. Stockholms Allmänna challenged the Bell Company by offering significantly lower subscription fees. Bell company responded by lowering their fees. The intense competition between the companies led to a rapid increase in the number of subscribers. In 1885, there were 5,000 telephone sets in Stockholm, which was more than in any other city in the world at that time! There also were around 400 other local networks all over Sweden at that time.

**A belated but rapid and entrepreneurial industrialization**

The liberal reforms, the freedom to start new companies, a political new order and free trade – which was introduced in the mid-1800s – set, along with investments in infrastructure and a new banking law, the foundations for a belated but rapid industrialization in Sweden. New manufacturers produced food, agricultural implements and textiles for the domestic market. The town Borås in the west of Sweden converted from a marketplace to a textile city and the town Norrköping became the centre for the wool industry (Schön 2008). The first steam sawmill began operating in Ådalen in northern Sweden in 1849 and was quickly followed by others. A huge international demand, especially from Britain, paved the way for a rapid expansion of the sawmill industry along the coast of Norrland.

Workers, without experience from their new profession, sought out from around the country to job opportunities in the new towns. They often lived in poor housing conditions, while new entrepreneurs could make huge profits and expose their new wealth in an extravagant way of living. But the market was characterized by rapid fluctuations in both demand and prices, which gave the whole industry a speculative character. Big profits and shortage of labour suddenly was replaced by bankruptcies and unemployment. In parallel many new successful export companies in the pulp and paper, steel and engineering industries were born during the second half of the 1800s. A new entrepreneurial capitalism flourished, which hardly resembled the traditional and stable social structure in Sweden.

The industrialization meant that people moved from rural areas to towns and cities as well as from agriculture to industry. But it was not a relocation of a constant number of citizens. During the 19th century Sweden’s population increased from just over 2 million to over 5 million despite a massive
emigration of more than one million Swedes, primarily to America. In 1850, there were still only 20,000 industrial workers across Sweden. But liberalization and communications speeded up the industrialization and created new jobs. It was among poor people on the countryside that the emerging industry recruited its workforce. Industrial work was, like emigration, an attractive alternative to a hopeless existence in poorest misery. It was mainly younger people who broke up and sought their livelihood in the new industries. They moved from place to place, and from company to company. They often ended up in the big cities that contained many companies and thus many alternative workplaces. To a large extent, it was also firms in cities who paid the highest salaries.

The emerging industrialization reformed society and created new social classes. A new working class emerged, which was free from the immediate and patriarchal control of their former landlords and masters who had compensated them in kind. Now they got their compensations by cash payment and they could sell their labour to the highest bidder on a free labour market. But the industrial freedom was limited in other ways. Work in the industry was regulated, disciplined, mechanized and monotonous. Working conditions were often poor, accidents common and working hours late. The financial and social security had also disappeared with the increased freedom. Shareholders, company directors and engineers, who ruled the industrial development, were often more anonymous, remote and inaccessible than the former masters.

Towards the end of the 1800s internationalism and free trade ideals did indeed give way to the rise of nationalism, but industrial manufacturing still pushed further momentum and increased prosperity. Exports tripled around the turn of the century, thanks mainly to the success of the engineering and the pulp and paper industries. But the increase in trade came to an abrupt end during World War I.

On August 1 1914 World War I broke out. The war had a great impact on Swedish industry even though Sweden stayed neutral and kept out of the war. Imports were strangled already in the beginning of the war, while exports continued for a few years. Sweden was transformed from a capital importing to a capital exporting country. Eventually, however, the Swedish neutrality policy also came to include the trade policy. As a result, Sweden’s foreign trade almost entirely ceased. Also trade with non-belligerent countries was complicated after 1917, when submarines started to attack trade vessels.

The loss of export markets hit many Swedish businesses hard. Most engineering companies however were richly compensated by an increasing domestic demand in an expanding war economy. They got large orders from government and some of them were extremely profitable during the war. Many profitable engineering companies also integrated backwards during the
war and increased their leverage through large investments in acquisitions, new machinery and the electrification of their fabrication. The investments were partly financed by loans. But a new Swedish bank act of 1911 also gave commercial banks the right to trade with shares. Many banks did that by starting new subsidiary investment companies, e.g. Investor. The banks combined their lending with buying new issued preference shares in borrowing companies. The new banking law also favoured a concentration of the banking industry in Sweden.

**Consolidation, centralization and corporatism**

An almost non-existing import during World War I created great problems for the Swedish population. There was a shortage of fuel, machinery, spare parts and consumables, but also of food. When the war broke out Sweden was far from self-sufficient in grain and feed. Both animal and plant foods became scarce. Rationing was introduced and ration coupons were handed out to Swedish citizens for sugar, coffee, butter, bread, potatoes, milk and so on. A continued population growth and urbanization also led to a significant shortage of housing in the cities.

The war in Europe also had great impact on domestic policies. The war became a source of inspiration for revolutionary forces all over Europe. In war-torn countries it created fertile ground for subversive revolutions. In Russia, the Tsar abdicated after the March revolution in 1917 and in August the same year the Bolsheviks took over power. Also in Germany revolts broke out and in 1918 Wilhelm II renounced his crown. The German Republic was proclaimed. The revolutionary currents also gained ground in Sweden. A general unrest in the country and numerous hunger demonstrations and riots supported the revolutionary agenda. The threat of a revolution appeared increasingly more real and paved the way for rapid and radical democratic reforms.

**Full democracy is introduced in Sweden**

Left winds were rash in the 1917 general election in Sweden. A liberal Prime Minister formed a new government and the Social Democrats participated for the very first time. Together they initiated a revision of the constitution, especially the introduction of universal suffrage for both men and women. In 1919 and 1921 the Swedish Parliament adopted the constitutional amendment which meant the democratic breakthrough in Sweden. In 1919 the Parliament also adopted a law on working hours. This meant that working hours were limited to 8 hours per day or 48 hours per week. The Swedish trade union movement had succeeded, with the help of government, and despite employers' unwillingness, to enforce the law. So the reformist worker’s union celebrated great triumphs. On the parliamentary road two of its main
demands, universal suffrage and the eight-hour day, had successfully been implemented. It was decisive victories and the revolutionary currents among workers in Sweden gradually declined.

During World War I and the years immediately following inflation rates had gone up. The prices quadrupled in 1913 to 1920. But the inflationary period was followed by a deep depression and a tough deflationary monetary policy, which led to prices again halved in two years. Companies that during the war had borrowed money from the banks and issued new preference shares ended up in difficulties. They still had their big debts, but the assets no longer corresponded to purchase values. They were also confronted with an extremely tough market. Many engineering companies, which had been favoured by the war economy, lacked customers when they would switch from military to civilian markets. Some Swedish companies also had assets in Russia, which were confiscated after the revolution.

Ownership in industry is concentrated to a few financial spheres

During 1920 to 1922 liquidations and bankruptcies succeeded each other in Swedish industry, which led to large transfers of ownership. Also many individuals were driven into bankruptcy and livelihoods deteriorated in the short term for a large majority of the population. Unemployment was widespread. In early 1922, a third of the members of LO (Sweden’s combined blue collar union) were unemployed. But from the rubble grew a new competitive export industry.

Many company founders; entrepreneurs, engineers and their families lost their businesses. But in came new owners; commercial banks and financiers. The number of commercial banks had peaked in 1908 and through mergers they were much fewer after the war. It was therefore a relatively small number of dominant ownership spheres within the Swedish financial circles who came to answer for the reconstruction. The control of the iron and steel industry, the pulp and paper industry and the engineering industry gathered at a small number of hands, which became even less after the Kreuger crash and a new depression in the beginning of the 1930s. In 1934 a very large part of the Swedish export industries was controlled by less than a dozen financial spheres.

The new principal owners of Swedish export companies started to implement consolidation and structural changes in the half-century old export industries. They were also eager to expand the markets and start rationalization and streamlining of operations in the companies that now belonged to them. They recruited rationalization oriented engineers as CEOs and other leading positions in engineering companies. The new executives were more industry strategists than former inventors and entrepreneurs. After World War I the profile of Swedish business leaders changed. You no longer became a CEO in
a large company by starting or inheriting a business, but by educating yourself and making a white collar career to the profession of a business leader.

Under the new leaders, many of the existing companies would develop very positively. However, after the mid-1930s industrial development was mostly about rationalization, restructuring and internationalization of existing businesses. More rarely, about creating or establishing new ventures. It was mainly within already established Swedish companies that one would come to pick up innovations and establish new product areas.

**Government increases its grip on infrastructure**

While the export industry consolidated and rationalized there was an apparent consolidation also in the infrastructure area. Government increased its grip on the infrastructure (Kaijser 2008:2). Government took, through the establishment of Vattenfall in 1909, first responsibility for regional power generation and later also the deployment of a nationwide transmission grid. Government took an even greater responsibility for the railways. In 1938 a decision was made on the nationalization of all private railways in Sweden. Swedish Rail (SJ) was then a national railway monopoly in the whole country. Telegrafverket successively acquired all local phone associations and telephone companies. In 1918, the acquisition was finally made of Stockholms Allmänna Telefonaktiebolag. Telegrafverket (later Televerket) thus created a nationwide de facto monopoly in the telephony field and turned to a Swedish PTT. From the 1930s the government did chose to direct and control the expansion and tariffs in a way that guaranteed companies and people in all parts of Sweden access to infra services on almost equal terms.

It was not only the control over the Swedish industry and infrastructure that was centralized in the 1930s. After the 1936 election the political power in Sweden was gathered in the reformist labour movement’s hands for many decades to come. The Social Democrats began, under the leadership of Per Albin Hansson, to build the Swedish welfare society. Already before 1936 Per Albin had developed very close, but discrete, personal contacts with some of the most important owners and CEOs in Swedish industry (Isaksson 2008). The Social Democratic vision included the creation of an efficient and streamlined Swedish export industry. The government also desired a peaceful labour market and the avoidance of blockades, boycotts and strikes. Efficient manufacturing processes and a competitive export industry was seen as the foundation of the Social Democratic vision for a future welfare society. It was the foundation that would allow for increased living standards for a large majority of the Swedish people.
A corporatist Swedish society is born

The interests of the new industrial owners and industry leaders' coincided largely with the reformist labour movement. After the election in 1936 the leading representatives of the large Swedish corporations chose to abandon their previous negative attitude towards the Social Democrats. Instead they decided to cooperate with the two branches of the reformist labour movement in Sweden; the worker's union (LO) and the government. They gathered around the common interest of rational and efficient export industries. After the 1936 LO Congress the reformist trade unions also took an active responsibility for the development of a more competitive Swedish industry and promoted rationalization; consolidation as well as automation and implementation of work study methods. Negotiations were started with the Employers' Confederation (SAF). A master treaty, the so-called “Saltsjöbaden Agreement”, was signed in 1938. The agreement gave the central organizations, SAF and LO, increased influence and acted in a clearly centralist direction.

The rapprochement between the central organizations on the labour market was followed by cooperation between government and industry. The outbreak of World War II in the fall of 1939 gave an extra boost to the corporate efforts in Sweden. Closures and refurbishment of defence also contributed to the close collaboration between government and parts of industry. Government introduced special military taxes and issued war loans, which allowed very large and rapid increases in public spending, largely to the Swedish engineering industry, whose advanced products gained an increasingly important role in the military defence. During the war, there was a pronounced spirit of cooperation both at the political level and between the parties on the labour market. Laws, including regulations about working hours and holidays, were breached and overtime and shift work became much more common. At the same time new regulations about official duty made strikes and conflicts impossible. During the war there also was in essence a wage freeze even in the expanding and lucrative engineering industry.

The good relations persisted after the war. The Prime Minister, Per Albin Hansson, who died in 1946, had replaced the socialist dogmatism for a role as pragmatic father of the nation. This was also true for his predecessor, Tage Erlander, who served as Prime Minister of Sweden and the leader for the Swedish Social Democratic Party from 1946 to 1969. The Social Democrats disassociated themselves from further nationalizations of private industry. World War II was, in contrast to World War I, followed by increased economic activity. Swedish industry was blessed with increased orders. During the 1950s and 1960s Swedish industry harvested major export successes. Sweden's share of world trade rose almost dramatically; Sweden took the lead
in terms of economic growth. The export industry increased its share of employment and was the driving force in the whole of the Swedish society.

The flourishing Swedish export industry laid the foundations of the welfare society that emerged in the decades after World War II. It created the preconditions in terms of economic growth and employment and contributed, both directly and indirectly, to the financing of the growing public sector. The success of the Swedish export industry was therefore not only a concern for business owners and leaders, but equally for the Social Democratic government and the labour movement. In the late 1960s, Sweden was one of the world's richest countries and the Swedish welfare society attracted international attention.

In a relatively short time a majority of the Swedish population had raised their standard of living from a precarious life of poverty to a life of full employment, relatively good income and seemingly great security. In less than half a century the average industrial worker had tripled their real wages. In addition, he was covered by pension insurance, medical insurance, unemployment insurance and statutory paid holiday. In the welfare society government also took responsibility for the welfare of citizens in terms of health care, education and housing. Government reallocated resources, through taxes and charges, between different stages of life, between rich and poor and between those who worked and those unemployed or unable to work.

There was a national consensus that laid the foundations of the welfare society. The development was driven systematically by central representatives of government, industry and trade unions in the context of a very centrally and informally managed truly corporative social society. The corporatist society was based on a clear division of roles between government, industry and trade unions. Some characteristics of the corporatist Swedish society are presented below.

**Efficiency in details and in overall structure**

The export success of Swedish industry after World War II laid the foundation for full employment. There was a shortage of labour – not a lack of work. Therefore, there was almost total agreement in Sweden that it is desirable to systematically streamline human labour in factories. The Swedish labour movement ensured that work and motion studies were made compulsory in Swedish industry in 1944 to ensure that objective piecework salaries were guaranteed throughout the engineering industry in Sweden (Giertz 1981).

The employers, the union movement and Social Democratic politicians all wanted to introduce a corporative version of Scientific Management. The most rational production methods were to be determined on objective grounds. A comprehensive change in the wage formation process for industrial workers
was one of the cornerstones underlying the carefully planned introduction of Scientific Management throughout the Swedish industry after World War II. In 1944 a verdict in the Labour Court (Arbetsdomstolen) stated that all piece rates in the engineering industry must be established objectively. All engineering companies were thus compelled to employ time-study engineers, to determine – on objective grounds – which working methods were to be applied for a certain task and the time required to perform that task.

As a consequence of the verdict the Employers Confederation and The Labour Union reached an agreement on work-study in 1948. Swedish companies thus not only had the right but a contractual obligation to employ motion and time study engineers and to perform time and motion studies on all manual work on the shop floors. Starting with the 1955 agreement the parties also agreed upon introducing a performance standard based upon Methods-Time Measurement (MTM). MTM is a predetermined motion time system to analyse the methods used to perform any manual operation or task and, as a product of that analysis, set the standard time in which an average worker should complete the task. MTM was launched by the US consultancy company Maynard’s in 1948 and soon became widely spread in the Swedish engineering industry. Thus, from 1955 the parties could check the performance appraisals and piece-work times made by any time and motion study engineer against hindsight. This absolutely unique arrangement is nowadays often overlooked in the description of the corporatist Swedish society which created the post-war economic miracle in Sweden.

The Labour Court’s verdict was entirely in line with the implications of the concept of the “policy of wage solidarity” and the principle of “equal pay for equal work” accepted at that time. Workers who performed the same volume of work were to be paid the same amount, regardless of which company or industry they were employed in. The ability of the individual company to pay, or the local supply of labour, was not to influence the earnings of the workers.

One outcome of the application of the principle of “equal pay for equal work” was – quite intentionally – that export companies in industries which internationally paid high wages, such as the automotive industry, were subsidized, while the rug was systematically pulled from under companies in industries that internationally paid low wages, such as the clothing industry. This type of guided structural change had been outlined in detail already in the 1930s by one of the leading ideologists of the Social Democratic Party, Sigfrid Hansson (brother of Per Albin Hansson). Labour was to be transferred from inefficient companies to efficient ones and from low-wage to high-wage industries. By means of retraining and “relocation policies” labour was to be transferred to those regions and sectors of industry that could afford to pay high wages. In this way, the value of the nation’s total production would increase, and people would generally enjoy higher living standards.
It was not only the labour productivity that came into focus, but also the mechanization and automation of production. The larger engineering corporations in Sweden generally coped very well with this transformation and they also took the lead when digital electronics were introduced in the factories. Corporations like ASEA, Volvo and Electrolux even started to develop industrial robots and auto carriers for internal use during the 1960s and later introduced them to the market. The Swedish machine tool industry also was very competitive and, until the 1970s, it was one of the key Swedish industries. It achieved international recognition, and as an equipment supplier to Swedish industry it played an important part in the success of Swedish manufacturing industry in general. You could definitely argue that production engineering was an important *developing block* (see Laestadius in this volume) in Swedish industry after WW II.

**Government facilitates mobility on the labour market**

In their deliberations in Saltsjöbaden in 1938, the parties on the labour market had declined governmental interference. To avoid governmental interference also became the main track for three decades in the post-war labour market policies. Government legislated working hours and holidays, but refrained in general from the regulation of employment conditions. In the agreement from 1938 the parties stated that it was the employers’ right to organize, manage and freely hire and fire employees (Giertz 1981). The agreement supported the structural changes and efficiency measures that the Social Democrats also sought. The government on the other hand, took an active role to facilitate labour mobility. It built up the resources to re-educate people, to move people and set up employment agencies. The relatively widely accepted Social Democratic “moving van policy” explicitly aimed at facilitating and supporting the restructuring and rationalization of the Swedish industry.

In the corporatist society the Social Democrats went arm-in-arm with the owners of major companies in Sweden. They pursued a policy that clearly favoured the accumulation of capital within established companies and disadvantaged new ventures (Jakobsson 2008). The list of listed companies at the Stockholm stock exchange also stayed more or less unchanged from the mid-1930s to the mid-1970s. At the same time the Social Democrats gradually tightened taxation of wealthy individuals through the introduction of progressive income tax and wealth tax and they increased gift and inheritance taxes. One consequence was that the dominance of the larger well established Swedish export companies increased heavily in Swedish industry. Ideologically the Social Democrats justified their close cooperation with the Swedish capitalists by pointing out that large established companies became increasingly dominant. It was argued that the individual entrepreneurship had played out its role. Consolidations and economy of scale would eventually lead to capitalism undermining its institutional and political importance.
Eventually it would peacefully pave the way for socialism in democratic countries.

**Demand-driven engineering sciences are introduced**

The corporative consensus also included an increased governmental interest in academic engineering research and education. During World War II government started to look upon engineering sciences as a national investment (Deiaco & Reitberger 2008). That thinking was inspired by Professor JD Bernal at the University of Cambridge. He argued that engineering sciences should be more demand-driven for the benefit of citizens, society and businesses. As a consequence, government decided that applied research, linked to common needs in different industries, should be performed by specific sectorial research institutes. Each institute would be co-funded by government along with incumbents in a specific industry. The research institutes would function as autonomous units but they were to be located next to corresponding institutions at the technical universities, especially KTH.

That a research program was jointly sanctioned by the large Swedish incumbents in one industry was sufficient justification for government to put in corresponding funding. Thus government completely relied on market forces – that of incumbents – to control the direction, expansion and contraction of applied research in various engineering fields. The set up created a major expansion and upsurge of the higher education and research in engineering sciences.

During the 1950s, there was a very common optimistic belief in a better future that relied on technological development. This belief was one reason to further increase governmental investments in Ph.D. studies and basic research at the technical universities. In parallel incumbents initiated applied research programs next door. The incumbents decided on large research projects, which were co-funded by government, in different sectorial research institutes. Sometimes the researchers almost didn’t know who of the colleagues were employed by the university and who belonged to an institute. This resulted in a Swedish model for technical universities that was quite unique. The institute sector stayed fairly small but there were strong links between basic research, applied research and industry. Education at technical universities was also closely connected to research and to actual problems in industry. This created a Swedish doctrine that was valid almost until the end of the 20th century.

The close cooperation and interplay between technical universities, research institutes and incumbents were combined with an extremely close cooperation between individual public customers and individual private corporations in Sweden. In those so-called developing couples, engineers and
technicians from public customers and private companies worked side by side with research, product development and implementation of new products and systems. This gave way for special conditions for the commercialization of research and new technology in the corporatist Swedish society.

**Developing couples paves the way for export industries**

As mentioned earlier, after World War I government assumed greater responsibility for the development and operation of the infra service systems in Sweden. The extension, expansion and modernization of the infra service systems involved major technical challenges and led to a very close relationship between public clients and the engineering corporations that supplied the technical input.

When power generation and the nationwide transmission grid was built it was the public company Vattenfall and the private company ASEA – the country’s leading manufacturer of electrical equipment – which jointly took on the challenge. They developed a close cooperation in high-voltage engineering. This form of close and long-term cooperation between a private Swedish company and a public customer has been named a developing couple (Fridlund 1998, 1999). ASEA obviously found it very valuable to constantly have a close cooperation with a demanding pilot customer. ASEA relied a lot on Vattenfall’s engineers who had great expertise in the operation of large power systems. Additionally, ASEA could perform pilot tests with new equipment in Vattenfall’s existing plants. It contributed greatly to ASEA in the 1950s becoming a world leader and a successful export company in high-voltage technology. Even in the nuclear sector ASEA got a leading position through public orders.

Developing couples played important roles in post-war Sweden. Each couple consisted of a public client and a private engineering company, which developed, manufactured and supplied necessary products and systems. Developing couples also arose within the railroad sector. SJ developed a close and long-term technical cooperation both with ASEA in the electrification of railroads and with LM Ericsson in developing signalling systems and equipment used for control and security of railway traffic. During the 1950s, an indigenous development of locomotives also started with ASEA that first resulted in thyristor converters for DC motor drives and Rc locomotives and later in asynchronous motors and X2000 trains (Kaijser 2008:2).

The most pronounced development couple, however, was the cooperation between the Swedish PTT, Televerket, and LM Ericsson. When Stockholms Allmänna was acquired by Televerket in 1918 many leading engineers and technicians in the company went over to their former supplier, LM Ericsson. Televerket and LM Ericsson soon initiated a close cooperation in the development of switchboards and other important components. The
collaboration between Ericsson and Televerket however differed somewhat from that in other developing couples. Televerket had since 1891 developed and manufactured their own products and systems in Teli, an industrial division of their own. Teli's first factory was located in Stockholm but was through a parliament decision moved to Nynäshamn in 1913. Later Teli established new factories all over Sweden in Vänersborg, Gothenburg, Sundsvall, Kristinehamn and Skellefteå. Since Televerket manufactured their own products they almost didn't buy anything from Ericsson. But Televerket and Ericsson joined forces to develop new products and systems.

The cooperation between the Swedish PTT, Televerket with its own industrial division Teli, and L M Ericsson was given an even firmer shape when they established a joint research and development company, Ellemtel Utvecklings AB. Ellemtel's primary task was to develop on its owners' behalf an electronic and automated telephone switching system for telephone stations that would become the AXE system. The new company was owned equally by Ericsson and Televerket. Ellemtel was purely a development company without its own production. Production would instead be the responsibility of the company's two owners using their respective production facilities in which Televerket and Ericsson manufactured more or less identical products for the domestic respectively the export markets. The close cooperation with Televerket gave L M Ericsson a great advantage in being able both to share development costs and to quickly gain experience from the operation of new products and components before they were launched abroad.

There were also other developing couples of great significance, such as The Air Force and FMV (The Materiel Administration of Swedish Armed Services) as clients and engineering companies like Saab as suppliers.

*More orthodox socialistic ideas rules government*

In the early 1970s the well-organized Swedish welfare state, with its high standard of living, still brought much attention and admiration abroad. But new forces were in motion under the idyllic and polished surface of the Swedish welfare society. The once docile inhabitants of the Swedish community started, in the late 1960s, to oppose those in power in the corporatist Swedish welfare society. A series of reactions against the established society was expressed. Protests, unrest and discontent also spread to the workplaces. In the early 1970s both contractual and unlawful strikes became increasingly common. The commotion contributed to the previously cherished work and motion studies and piece rate systems being thrown out of the Swedish industry. Socio-technical systems and autonomous groups was instead the new mantra regarding work organization in Swedish factories. These changes occurred at the same time as demands for real socialization received backing from a new generation of leading Social Democrats.
More orthodox socialistic ideas took rote in government in 1969 when Olof Palme succeeded Tage Erlander as Prime Minister in Sweden. Palme rejuvenated the government, paving the way for more radical ideas which had taken root among the young Social Democrats. A few years earlier, in 1967, the Social Democratic Party Youth League (SSU) held a Congress that really reflected the left winds (Ljunggren 2008). The Congress adopted a very radical action plan. The resigning President, the future Prime Minister Ingvar Carlsson, and the new President elected by the Congress, the future Minister of Finance Bo Ringholm, had endorsed the congressional opinions. The Congress demanded, among other things, the nationalization of the banking industry, pharmaceutical industry, pharmacy services, construction materials industry and petrol and oil trading. SSU also demanded an end to the cooperation with capitalists and demanded that citizens would have real power and control over the means of production. Thus the government’s close collaboration with the capitalists should be replaced with more orthodox socialistic ideas.

Krister Wickman took up a new position as Minister of Industry in Olof Palme’s cabinet. Already May 29, 1969, he presented Bill 121, which became the basis for the formation of Statsföretag AB; an industrial conglomerate with extravagant expansion plans (Fredriksson 2008). The idea was that government, through Statsföretag, would be responsible for the production of basic goods in the Swedish society, such as schoolbooks, computers for pupils, plastic bicycles, postal cars, steel, food and medicines.

Government also introduced a more active industrial policy in other respects. A new governmental agency, the Board for Technical Development (STU) was established in 1969. Government argued that the innovation process – the transfer of ideas and research results into commercially successful products – must be seen as a single, cohesive chain of activities. When STU was established government also started a new investment company. One mission was to exploit innovations made by employees in government authorities and agencies and state-owned companies. Formally the state's more passive role in the industrial development would be replaced by an active, proactive involvement at government level. It meant a sharp departure from the previously established division of roles between government and industry, where government completely handed over the commercialization to private incumbents.

While STU was awarded a role in the development and commercialization of new technology a more hostile approach towards rationalization and new technology also gained a foothold among social democrats and in the trade unions. The increasing computerization, automation and use of industrial robots were particularly highlighted. Demands were raised to stop or stem the use of new technology, so that people had enough time to adapt to new conditions. Government started new delegations (Dataeffektutredningen,
Datadelegationen and Data- och elektronikkommittén). This signalled a very radical change of perspective in the Swedish labour movement that spread real fast.

Even if few people noticed, you could in the early 1970s also see the first signals of an upcoming crisis. As a consequence, subsidizes became a part of the newly established concept of industrial policy. Selective support measures, targeted at industries dominated by small enterprises, like textile and clothing, footwear, glass, furniture and foundry industry, were the result. The responsibility for handling the programs of selective subsidizes was given to a new government agency, the National Industrial Works (SIND), which was established in 1973. During the period 1973-75 pay-outs doubled, mainly due to increased support operations in crisis-affected companies and sectors, but also through increased R&D funding and increased payments for export promotion.

**Wage-earner funds initiates great tensions**

Despite the “oil crisis”, and the increased costs of industrial policy, there was not a widespread “sense of crisis” in the country before 1976 (this section draws on Giertz (1991) p. 253-269 and 387-407). In 1974, the major export companies made very high profits overall. The political debate was dominated by discussions about the “excess profits” – not on “industrial crisis”. Growth and increased standards of living were seen rather as governed by the laws of nature than as a result of innovations, rationalization and hard work of the population. The older social democrats' missionary on these matters was more or less forgotten, which gave significant consequences for the actions taken by government and trade unions.

Sweden’s confederation of blue-collar unions (LO) pushed the issue of economic democracy at the beginning of the 1970s. The list of companies on the Swedish stock exchange had remained basically intact since the mid-1930s. To own shares therefore appeared as safe as putting money in the bank, and there had almost been no need for issuing new shares or for venture capital as new technology, new products and new markets had been intercepted by the established incumbents. In 1971, the convention of LO proposed a study to investigate the possibility of introducing collective capital formation. In 1975, the study, under the leadership of LO economist Rudolf Meidner, was presented. “The Meidner Plan” recommended the introduction of wage-earner funds. It was proposed that 20 percent of a company’s profits should be used to buy newly issued equity shares. This proposal entailed a gradual transfer of ownership from private individuals and institutions to collective entities governed by union-appointed directors. The funds would eventually own 50 percent of companies in the long term. The LO convention supported and approved the Meidner Plan in the summer of 1976.
The discussions about wage-earner funds were partly ditched when the Social Democrats lost the election in the autumn of 1976. But in 1982 Olof Palme was forced, partly against his will, to make wage-earner funds one of the election’s key issues. LO demanded that wage-earner funds be given a prominent position on the Social Democrats’ agenda. However, a considerably modified proposal had been put on the table by that time. A group led by the future Minister of Finance Kjell-Olof Feldt, had revised the original plan. When the Social Democrats won the election wage-earner funds were introduced in a modified version in 1983. However, the taxation was immediately cancelled in 1991 when a right winged government, under the Conservative Prime Minister Carl Bildt’s leadership, was appointed and a few years later the funds were closed down.

When the ideas of wage-earner funds started to take shape in the beginning of the 1970s, it was the death knell for the government’s earlier cooperation with industry. The tension between industrial owners and CEOs on the one hand and leading Social Democrats and labour union leaders on the other increased considerably at the beginning of the 1970s. Thus the former foundation of the corporate Swedish society was destroyed.

Informal cooperation is replaced by formal bureaucracy

The tension between government and industry increased even more when government abandoned its former non-interventionist role on the labour market. Investment funds, which encouraged companies to reduce tax by locating manufacturing plants to regions with unemployment, were introduced. The union movement also relinquished its tradition of negotiating and agreeing with employers on numerous common issues, and instead turned to government with a request for legislation. Various labour law reforms, including the Act concerning Security of Employment, Board Representation for the Employees, Position of Union Representatives and Co-determination at Work, were implemented by the Minister of Justice, Carl Lidbom, during the period 1972-77. The new labour laws resulted in an almost complete turnaround compared to the previous “Spirit of Saltsjöbaden”. There is surprisingly little attention paid to this turnaround in more recent discussions about the content of the so called Swedish model. There might even be some people who believe that the 1970s labour law reforms are part of the original Swedish model and a reason for the Swedish economic miracle in the 1960s.

Behind the visible changes also a subtler but important development lurked. The relationship between ministers in the government and the business community’s top representative changed radically. In previous Social Democratic post-war governments, the Prime Minister as well as other important ministers – such as the Ministers of Trade and Finance – had close personal contacts with top representatives in important financial spheres and
with employed CEOs. In the 1970s however, confidentiality, personal contacts and informal consultations between government and business leaders were replaced by bureaucratic contacts between officials, officers and ombudsmen of various kinds. The direct contacts between government and industry became extremely frosty. The earlier consensus had been replaced by powerful confrontation at crucial points. In parallel the public sector developed its own staff organization in the 1970s. Government agencies, public authorities, ministers and organizations representing various groups in society took over the handling of an increasing number of tasks.

The personal contacts between politicians and industry leaders were succeeded by interaction between civil servants from the large public authorities, leading trade unions and large employer associations. Oddly enough these stiff structures were put in place when Swedish industry was just about to enter into a new phase characterized by great industrial dynamics.

**From crises to an entrepreneurial global economy**

In 1976 the Social Democrats lost the general election. After 44 years of uninterrupted government a liberal-conservative coalition, led by the Centre Party leader Thorbjörn Fälldin took over. His predecessor said that when entering the cabinet, the new government could just start picking presents off a Christmas tree arranged by the Social Democrats. But the new Minister of Industry, Nils G Åsling, barely had time to take up the post until his office became an Emergency Room for failing firms. Sweden’s economy was greatly unbalanced. A shrinking export industry could no longer finance government’s galloping spending.

Shortly after the seemingly good “excess profit year” in 1974 – which also brought greatly increased wage costs for industry – structural problems appeared. Productivity decreased and real earnings were reduced. Unemployment rose, particularly among young people. Special support measures were directed at small enterprises and companies who arranged jobs for young people received special grants. But above all there was a dramatic increase in direct efforts to sustain employment in the country’s ailing basic industries and incumbents. Concealed behind the increase was direct aid to shipbuilding, steel, forestry, mining and textile industry. During the period 1977-79 over nine billion crowns was paid in direct subsidies only to the shipbuilding industry, which represented approximately 280,000 crowns per employee, or 120 percent of the shipbuilding industry’s total payroll. The crisis also made, ironically, the new liberal-conservative government responsible for the largest socializations in the country’s history. With an ambition to pursue structural rationalization, government entered as
the sole owner of several companies; for example, Svenskt Stål AB (SSAB) and Svenska Varv AB (Swedish Shipyards).

**SME:s and Universities in focus in regional policy**

When the incumbents in basic industries went into trouble the country's total production was greatly reduced. The industrial crisis also spoke quickly in reduced market shares on the world market, rising unemployment, negative trade balance and low profitability in many industries. In the wake of the crisis the liberal-conservative government's interest to encourage the growth of small and medium sized enterprises increased. This also initiated a major change in the institutional conditions that set the scene for the Swedish innovation system (see Rickne and Laestadius in this volume).

As a result, the business associations were converted to regional Development Funds in 1978. SIND and the county council in each county became principals. The Development Funds, which received increased resources and access to their own loan funds, would primarily focus their support to start-ups and existing businesses with up to 200 employees.

The election in 1982 reiterated the Social Democrats to power. The crisis had drained government finances. A widespread sense of crisis had helped to keep the Swedish wage increases and inflation in line with the main competitor countries. Consolidations and eliminations had also cleaned up the structure in industry. The remaining companies did well, and the industry's competitiveness on foreign markets was further improved by a devaluation of sixteen percent, which the Minister of Finance Kjell-Olof Feldt pushed through immediately when the new government was installed in 1982.

Swedish industry in general started to recover in the 1980s but there were significant regional imbalances. The eliminations of plants, enterprises and certain industries had resulted in a redistribution of employment. Employment increased in those cities where a dozen successful engineering corporations or pharmaceutical groups decided to expand, but declined sharply in regions with previous employment in industries such as shipbuilding and steel works. The regional imbalances had great impact on industrial policy in the mid-1980s, when Ingvar Carlsson assumed the position of Prime Minister after the assassination of Olof Palme.

In the mid-1980s the Social Democratic Minister of Industry, Thage G. Peterson, no longer had the possibility to direct an increasing number of blue collar jobs to regions with unemployment. Now when successful Swedish industrial corporations gained market share, invested in research and development, introduced new technology and improved their profitability, they did not increase the number of blue-collar jobs in Swedish factories. On the contrary, manufacturing industry's share of employment in Sweden
decreased throughout the 1980s. Nearly 100,000 jobs disappeared from the export industry. At the end of the decade, all companies producing any kind of goods accounted for less than 20 percent of employment on the Swedish labour market.

While routine factory jobs vanished, the need for skilled technicians and engineers increased in the export industry. Overall, the industry doubled its spending on research and development during the 1980s. The need for MSc’s in engineering took off. But the investments were concentrated to a little more than a dozen corporations in the engineering, pharmaceutical and chemical industries. They concentrated largely on the metropolitan areas and the vicinity of universities. A lot of investigations, as well as the media, also highlighted research as a vehicle for growth in the emerging knowledge society. Cities with a university were pinpointed as the winners in the future society (Andersson 1988). They argued that there was strong links between research and higher education, the creation of new competence, the growth of new knowledge-based businesses and thus new jobs (Deiaco 2002). Cities with universities were pinpointed as the growth centres and the labour markets of the future, where highly skilled, highly paid and creative people flocked.

Media made out the universities to be the future engines for the growth of new knowledge-based companies. For regional policy-makers it seemed obvious that there were strong links between research, new knowledge-based companies and future employment. International studies also supported a positive correlation between the percentage of highly educated and the economic growth of a country. This made higher mass-education get its definitive breakthrough in Sweden. The expansion also enabled the dispersion of both higher education and research.

In 1988, Sweden received four new universities in Luleå, Karlstad, Växjö and Örebro. At the same time there were investments made in regional colleges, which almost invariably came to engage themselves in some kind of research. Thus from the mid-1980s research and higher education became an instrument of regional policy. New universities and regional colleges were the base for regional business policies. They were not only designed to satisfy the need for skilled labour on the regional market. They would also conduct their own research and hopefully become the breeding ground for comprehensive new enterprises with great employment potential.

*Internationalization and decreased productivity hits manufacturing*

The end of the 1980s was characterized by heightened global trade. Countries that previously had very protectionist trade policies joined forces in large trading blocs that positioned themselves in the global arena. This change was most evident in Western Europe, which strived to resuscitate its former role
as a great industrial power. In 1986, after Spain and Portugal joined the
European Economic Community (EEC), the organization comprised 12
member nations and the process had begun to convert it into a type of
European federation. Everything would be eliminated that hindered free
movement between member countries for goods, services, capital and people.
By the end of the 1980s, this vision of creating a single internal market had
started to take concrete shape.

When the decade shifted in 1989-90, developments in Europe took yet
another turn. The walls between East and West were torn down at an
astonishing speed. On October 3, 1990, the two German states reunited. The
Soviet Union was dissolved and Eastern European countries became
independent nations that embarked on a difficult journey toward democracy
and a market economy.

For the nation of Sweden, many of its curves started to bend in the wrong
direction as of 1987. On an international scale, Sweden still had an extremely
low unemployment rate, which was very much due to the expansion of the
public sector and the growth of the automotive industry. Wages and salaries
soared to new heights as a consequence. In combination with price increases,
for which insufficient capacity utilization within industry was another cause,
the competitiveness of Swedish industry declined even more. Market shares
for Sweden’s engineering industry decreased 10 percent in three years and
Sweden’s balance of payments plunged radically. Companies were forced to
borrow more money outside the country and interest rates skyrocketed. The
trade surplus also started to decline, despite a strong economy, low oil prices
and a falling exchange rate for the U.S. dollar, which had a devaluation effect
on the Swedish krona in the most important European export markets.

The new economic reality caught the eye of the political establishment towards
the end of the 1980s. The Social Democratic government in Sweden, under the
leadership of Ingvar Carlsson, carried out major reforms to Swedish economic
policy. The Minister of Finance, Kjell-Olof Feldt, was the driving force behind
many important changes that enabled Sweden to adapt better to the rapid
internationalization of its economy. In a short time, a number of measures
were carried out, including deregulation of the credit market and abolishment
of foreign exchange controls.

The engineering industry in Sweden was also becoming internationalized. In
the beginning of the 1980s, it was a Swedish export industry with the majority
of its employees in Sweden. By the end of the decade, it was an international
industry with most personnel abroad. Thus, at the close of the 1980s, major
Swedish engineering companies had expanded their foreign production base,
due in part to acquisitions of international competitors, but also due to
substantial direct investments in other countries at the end of the decade.
They rationalized their international expansion with arguments that they
needed to get closer to the market or have local production in countries, but in actual fact, Swedish production units had become less profitable in the 1970s and the 1980s, and the corporate leaders of major Swedish corporations like Ericsson, Electrolux, Atlas Copco and Sandvik explained, quite frankly, that Sweden was less and less of an economically viable alternative, compared to the possibility of setting up operations locally in other countries, which they were contemplating. The reasons were the country’s high wage costs, shortage of labour, high sick leave rates and low productivity levels. At the end of the 1980s, many Swedish engineering companies invested directly within the tariff walls of the EEC in countries like Germany, France, the Netherlands, Italy and Spain. In the early 1990s, labour-intensive production was moved out of Sweden also to low-cost countries in Eastern Europe.

**New crises set the scene for a more competitive future**

In the beginning of the 1990s a recession was waiting. It was magnified by Iraq’s invasion of Kuwait in the summer of 1990. The industry was forced to adjust production capacity to a declining demand. Cutbacks arose in virtually all industries. Questions concerning the location of production no longer dealt with the location of new plants but rather where necessary closures would be made. Confidence that incumbents could guarantee full employment in the future faltered in almost all political parties and on all levels. In the industrial and regional policy more and more weight was given to new ventures and to small and midsize enterprises. This also enforced a shift in the governmental regional policy, from measures that primarily reallocated employment, to efforts that supported regional or local development programs. By now the Social Democrats were no longer very keen on owning the Swedish industry. The industrial dynamics during the 1970s and 1980s, and the failures in state-owned companies, made politicians in almost all parties hesitate towards governmental ownership in industry.

In the wake of the recession prices on commercial real estate plummeted. In two years they fell by more than 50 percent. New construction ceased, which hit hard on the construction and building industries. The downturn on the property market first hit the real estate companies, but also had serious repercussions on investment companies and banks. The Social Democratic government presented a “crisis package”, which included savings in the public sector and reduced health insurance. The Social Democrats also made changes related to industrial policy. Three governmental agencies – SIND, STU and Statens energiverk – merged into a new governmental agency, The Swedish National Board for Industrial and Technical Development (Närings- och teknikutvecklingsverket, NUTEK).

In 1991 a very radical tax reform, agreed upon two years earlier both by the Social Democrats and the liberal-conservative parties, was implemented. It lowered the previous extreme progressivity of the Swedish income tax to more
normal levels. It was still a progressive income tax, but in principle, the Swede would in the future not pay more than half of a wage increase or an extra income in tax. The progressive income tax was combined with a flat tax, of 30 percent, on capital gains.

In the chaotic situation, the Social Democrats once again left government. After the 1991 election conservative leader Carl Bildt became Prime Minister of a liberal-conservative government. The financial crisis forced a series of costly commitments. Alongside dealing with the unemployment, government had to guarantee bank liabilities to third parties, give subsidies to support some banks and also nationalize the private bank Nordbanken. In May 1993 government set up a special governmental agency, Bankstödsnämnden, to handle further support to banks in trouble. The crisis weakened public finances and unemployment was widespread. Nobody put trust in Swedish economy any longer which imposed new problems. When the credit market was deregulated the Swedish krona had been linked to the European currency unit ECU at a fixed exchange rate. The Riksbank – Sweden’s central bank – tried the longest to stave off the attacks against the krona. The overnight interest rate even rose to improbable 500 percent on September 16, 1992, before the Governor, Bengt Dennis, gave up. The krona was allowed to float, which in reality meant that it sank like a stone in November 1992.

**Joining EU facilitates the consolidation of Swedish economy**

From the beginning of the 1990s Sweden’s relationship to the European Community was an important issue in the growth debate in Sweden. As the vision of a common internal market took concrete form, it became increasingly clear that Sweden would join the European Union (EU). EU was realized in November 1993, and in January 1995 it was extended to include Finland, Sweden and Austria. EU’s so-called convergence criteria facilitated the consolidation of the Swedish economy during the first half of 1990’s.

When Sweden joined the European Union the Social Democrats had returned to government. In 1994 Ingvar Carlsson put together his third Cabinet. Göran Persson was appointed Minister of Finance. He implemented welfare cuts and tax increases to place Sweden in a position to qualify for the European Economic and Monetary Union. The job was done and the budget deficit went down but unemployment rose considerably.

International confidence in the Swedish economy was restored. The krona was appreciated and interest rates were reduced to historically low levels. Inflation ceased almost completely, but a larger decline in the unemployment rate was slow in coming. The development in Europe also took a different turn after the walls between East and the West were torn down. Over time it created a greatly enlarged EU.
Deregulation, privatization and productivity on the political agenda

In Great Britain the Conservative government of Margaret Thatcher started a program of deregulation and privatization in the financial sector, transport sector and the telecom sector in the beginning of the 1980s. This liberalization movement spread to other countries.

Already in the end of the 1980s the so-called right-wing in in the Swedish Social Democratic government, under the leadership of Kjell-Olof Feldt, had secured that productivity issues were once again put high on the social democratic agenda. Feldt spoke as former social democrats did, when he stressed that growth requires a more efficient production throughout society – of goods as well as services. This could be the result of introducing new technologies, but also from introducing new and more efficient regulations, management, structures and work methods in working life.

A productivity delegation was appointed in 1989. It stressed that real growth is based on efficiency improvements, which ensure that we get better productivity and quality and better service in all parts of working life. The delegation’s report, which was delivered in 1991 (SOU 1991:82), focused on weak productivity improvements in the Swedish working life. The delegation looked at the nation’s total production capacity and stressed that the transformation pressure, especially in service sectors, must increase.

A wave of deregulations in sectors such as transport, electricity, media, postal services and telecommunications were rolled out. Sweden was to play a leading role in the transformation when firms and agencies in domestic monopoly or oligopoly industries faced competition from new competitors and international players who expanded their geographic markets.

The taxi sector was deregulated in 1990. The reform aimed to improve access to taxis at all hours and across the country, by making it easier for taxi drivers to set up a business. By increasing competition, it was also expected that taxi fares would stay at a relatively low level. As a result, traditional federative taxi networks were challenged by new franchising networks. In order to increase competition and lower prices, domestic air travel market regulations changed in 1992. The reform meant that former privileged airlines, like SAS and Linjeflyg, met new competition from budget airlines. The railroad monopoly however stayed untouched for long. In 1988 a cut was made through the nationalized railway monopoly (SJ); since then, infrastructure is handled by a public-sector agency (Banverket) while trains until 2011 were run by the government-owned monopolist (SJ). Since 2011 however SJ competes with other railway companies.

A deregulation of the Swedish electricity industry was started in the 1990s. A first step in the process was the corporatization of the Swedish state-owned
utility Vattenfall in 1992. The deregulatory process culminated with a new Electricity Law, which entered into force in 1996 (Högselius and Kaijser 2007). The new regulation introduced competition on the electricity market in the generation and distribution of electricity. The aim was to provide consumers with greater freedom of choice and better opportunities for putting pressure on costs and prices in the supply of electricity. When charging customers, the total cost of electricity should be separated into two components, the transmission price and the price of the electricity.

Deregulations and privatizations of different state-owned monopolies have continued over time. In 2009 Sweden’s state-owned pharmacy monopoly, which was originally formed in 1969, ended. A huge part of about 1,000 individual pharmacies were sold to large and medium-sized companies in eight different clusters. But the deregulation also paved the way for new entrants into the market for drug sales. A similar procedure followed in 2010 when the market for annual, compulsory vehicle control was opened for competition. Previously all vehicle controls had been provided by a national monopolist. The government has been partitioning the incumbent and gradually selling it on commercial terms.

In addition, Swedish governments have, since the 1990s, experimented with boosting public-service efficiency in different sectors by allowing private actors to compete with publicly owned vendors on tax-financed markets. Many schools are now independently run, and in health care as well as child and elderly care private management is a growing – and frequently debated – trend.

But you can definitely argue that everything must not only be left to market forces. For instance, Rurik Holmberg argues that the potential for energy efficiency will not reached if it is not embedded in a broader societal context (see Holmberg in this volume).

Outsourcing shapes new specialized industries

The deregulation and privatization in the public sector was accompanied by a similar trend in many industries. Already in the mid-1980s many Swedish incumbents started to concentrate on their core activities. They spun off numerous businesses and got other companies to supply everything that was not regarded as their own core competence.

Outsourcing, the hiving off of activities that are considered peripheral, did affect simple activities just as much as advanced or complex ones. The activities that an incumbent decided not to keep in-house were sourced from suppliers, with whom they often have long-standing relationships. Their own employees were in that way replaced by a more flexible and specialized external work force, with both highly standardized and highly specialized
services. The external work force may be self-employed consultants, employees of large consultancy corporations, call centres, facility management providers or agencies specializing in supplying anything from doctors to blue-collar workers. Also the manufacturing of components was outsourced in many engineering companies. An example was Ericsson, who in 1996 decided to sell production plants in Sweden – manufacturing components they still needed – to contract manufacturers.

For the outsourcing companies, concentrating on their own core activities and selling off other activities caused major changes in the relationships they had with those involved in their business systems. Spinning off numerous employees into different independent companies of course resulted in fewer employment contracts and more business contracts for the outsourcing company. Moreover, the new business contracts sometimes involve a stronger tie than the former employment contracts, since the latter could be cancelled by the employee, while a company which has a contractual obligation to deliver must fulfil its undertaking in order to retain its credibility on the market.

![Diagram](image)

**Figure 1:** In the mid-1980s large incumbents started to spin off different activities, from fairly simple local services to professional services, manufacturing, computer operations to sales activities. In the beginning they sometimes formed new companies, called for instance Support AB, Assist AB, Partner AB, Consultancy AB, Education AB or Data AB, within their own corporation. Eventually many of the new companies were later acquired by specialized service companies in different service sectors.
The outsourcing of different staff and service activities in private as well as public sector created expanding B2B industries of different kind. Some expanding industries offer customers services on a local market, i.e. repair and maintenance, janitorial services, installations, painting services, security services, facility management, local transportations and canteen services. Others offer more knowledge-intensive professional services, i.e. accountants, technical consultants, IT consultants, architects, management consultants and marketing agencies. Due to technical development some support could also be provided on a distance. New work places providing distance support, i.e. computer operations monitoring, call centres and help desks, were in the 1990s, by new vendors, concentrated to locations where suitable labour was available. Finally, a lot of large companies made changes in their staffing of sales activities in the beginning of the 1990s. Employed sales forces appointed for an indefinite period were complemented by a new kind of sales forces, i.e. franchisers, telemarketing and sales agencies. For the new sales force, sales commission accounted for the lion´s share of revenues.

The outsourcing of different services challenged the established system for wage formations in Sweden. The traditional system was built on wage formation at central level between the parties on the labour market. It also included “the industrial union principle”, meaning that all blue-collar or white-collar employees in an industry were covered by the same agreement, but the agreements in another industry could differ. As a consequence, a driver in a manufacturing company could earn more, especially when driving at night, than a driver employed by a transportation company. These differences accelerated the outsourcing of some activities.

When a former staff activity was divested from an outsourcing company it found itself in a commercial environment where it was exposed for competition. This also involved a change in conditions for the employees, whose wages were influenced to a greater degree by the company’s profit level and the competition on the local labour market. It also meant that what used to be a peripheral and perhaps neglected part of a large incumbent suddenly was transformed into a core activity in a spin-off company.

Wage formations and earnings are nowadays very different in different parts of the service industries in Sweden. For large groups of employees, for example in the manufacturing industry and the public sector, wages are still controlled in a more traditional manner. Wage differences between different professions and positions are relatively small, and wages are almost completely independent of individual performance or results. Other industries, i.e. janitors and call centres, are still controlled by central agreements but wage floor has been lowered compared to manufacturing industry and different kinds of incentives have been introduced. In other industries and activities, i.e. consultancy and sales, wage formation is controlled by market forces outside the central parties control. Wages are
relatively high but wage differences are also great. In many of those industries, there is a clear link between the earnings and the individual employee's job performance and the revenues of the individual work place or company.

**Televerket and Ericsson are divorced**

The deregulation, internationalization and privatization broke up most developing couples in Sweden. However, the very close cooperation between the monopoly Televerket and Ericsson continued until the mid-1990s (Giertz 2015:1). In fact, a gigantic development project was started in 1987 in the jointly owned development company Ellemtel, which had previously developed the very successful AXE system. The new project was named AXE-N, where N stood for Network. But the benefits from a successful project would be substituted by internet that simultaneously began to take shape in the early 1990s. The AXE-N project was stalled in 1994 and completely closed down on December 8, 1995. This was a chock not only for 700 employees but also for a number of R&D related ICT consultancy companies who almost overnight were out of job.

Towards the end of the project the marriage between Televerket and Ericsson ended up in a divorce. They had been a developing couple for about 70 years but the time to separate had come. Already in 1990 government had taken some steps on a liberalization journey by transforming Televerket into a public company. In 1993 the new company Telia AB was formed and a new Telecommunications Act was launched. This initiated a first significant liberalization effect. When Telia was to become a competitor in the telecommunication industry and not a state-owned domestic monopolist a lot of things changed. In 1994 Ericsson bought Telia's equipment-manufacturing units, which meant that the domestic AXE manufacturing was transferred to Ericsson. In 1995 Ericsson also bought Telia's holding in Ellemtel.

On June 13, 2000 a partial privatization of Telia AB, through an initial public offering (IPO) on the Stockholm stock exchange, was realized. It was a huge interest among Swedes in general. Government decided that all investors who applied for 100 shares or 400 shares would get them, while those who applied for more only got a fraction. The Minister of Industry and Infrastructure, Björn Rosengren, made adds about the "people's share" and 954,000 Swedes actually became shareholders in Telia AB. The IPO was done just before the third generation mobile standard, 3G or UMTS (Universal Mobile Telecommunications System), was about to be installed in Europe and Japan. Licenses for the new system were issued in the European countries in the late 1990s and early 2000s. Different countries chose different approaches.

The Swedish Post and Telecom Authority (PTS) chose to allocate the licenses through a “beauty contest”. Telia applied but on December 16, 2000 PTS
announced that Telia did not receive a license. Telia was the only former PTT in the world finding itself in this position. The licenses were instead allocated to Europolitan (which later became Vodafone and then Telenor), Hi3G (later Orange (later acquired by France Telecom) and Tele2. Telia was in shock. The former monopolist not only met new competitors. Telia was not even allowed to compete on the mobile market. But Telia found a solution. Telia and its competitor Tele2 soon announced their intention to create a joint network company, Swedish UMTS Network Company, to utilize the 3G license which Tele2 had been awarded. It was a joint venture on a 50-50 basis, which was approved by the Competition Authority in the spring of 2002.

**Ericsson survived when the bubble busted**

In the year 2000 Ericsson was riding high (Giertz 2015:1). Ericsson had more than 100,000 employees globally, despite the fact that Ericsson had closed down their own production of processors and other components and outsourced manufacturing plants to Flextronic in the late 1990s. Still Ericsson had more than 43,000 employees in Sweden. The skills and competence of Swedish employees however, had changed over time. Blue collar workers had been replaced by engineers and technicians of various kinds.

In the year 2000 the turnover in Ericsson was close to 300 Bsek and Ericsson’s share of Swedish GDP was two percent. Ericsson was in the very centre of a worldwide hype around the potential of the internet. Ericsson had around 200 million AXE lines in place in 120 countries. But the success was very much linked to a fast penetration of GSM systems in the late 1990s. Ericsson held an estimated 40 percent share of the world’s mobile market. At the turn of the millennium Ericsson was on top and expecting to deliver a lot of brand new 3G-systems, especially to tele operators in Europe. But the market failed!

One important reason why the market failed was that most European countries chose to allocate the 3G licenses through auctions. European tele operators spent around 130 billion Euro to buy licenses for 3G-networks that were supposed to give people the freedom to use their mobile phones for reading e-mail, browsing the Internet, placing video calls, enjoying music, buying products and services, making reservations and so on. The nature of the auctions was designed to increase competitive pressure on bidders by offering fewer licenses than the number of operators likely to bid. The tele operators were put in a difficult position, because if they lost the auction they were out of business in the next phase. They took risks and made high bids and ended up with large debts. But when they got the licenses they could not afford to buy and install the systems.

The tele operators went into serious trouble. When the trouble leaked out to the public the hype on the stock market was put to an end. On March 6, 2000 the index on the Stockholm stock exchange had been all time high. The total
value was 4,800 Bsek and Ericsson alone was valued to 1,709 BSek. The next
day the bubble started to burst, but Ericsson still showed good figures in 2000
and coped rather well initially. The share value of tele operators on the other
hand fell quickly all over Europe. They had no choice but to reduce their
investments sharply. This of course also hit Ericsson and on October 20, 2000
the share price went down drastically. Still the figures for the full fiscal year
2000 were quite ok.

In Ericsson the crisis was a fact in March 2001 (Karlsson & Lugn 2009). More
than 700,000 employees in the tele vendor industry around the world lost
their jobs in a very short time. There also were a lot of consolidations
throughout the industry. But people still wanted new mobile phones, so the
market for handsets was almost not affected at all, which partly explains why
Nokia was not hit that hard. Ericsson on the other hand had great problems
with their handsets. Former, when the demand exceeded the supply, Ericsson
had done well because their phones were reliable. Ericsson still shared the top
position with Nokia and Motorola in the late 1990s. But in 2000 people on the
consumer market started to ask for more. Ericsson decided to concentrate on
core business, that is tele system – not consumer products. Despite cost
cuttings and massive layoffs, the fiscal year 2001 ended with a very big loss. It
was Ericsson’s first loss in more than half a century. More than 8,000
consultants also lost their customer in 2001. But it was not only a question of
profit and loss, but also about the cash flow situation. The lack of orders from
the tele operators was one thing but customers also paid late due to bad
liquidity. Ericsson was forced to borrow money to pay salaries. The situation
was alarming in March 2002. The customers were afraid to put orders because
they did not believe Ericsson would survive. The board decided to issue new
shares. Ericsson was only weeks from bankruptcy when the new money turned
up.

Ericsson survived and took off again in 2004. By then global employment was
down to a little more than 50,000 employees globally and 21,000 in Sweden.
Thus Ericsson’s global and Swedish employment fell to less than half of their
previous levels in only four years. In Sweden, the ICT sector outside Ericsson
was also hit. Via its external networking and partnering, Ericsson had
contracted tens of thousands of Swedish companies in ICT-related
development, consulting, manufacturing and maintenance. In the successful
attempt to return to profitability, Ericsson slashed this spending. With this
multiplicative effect and Ericsson’s own layoffs, a huge volume of ICT-
expertise was re-deployed in Sweden. And that is still the case. Ericsson’s
employment worldwide has increased since 2004, but employment in Sweden
is continuously decreasing. Ericsson is no longer only a manufacturing
company exporting goods, but also a global service provider that operates
telecom system worldwide on a contractual basis.
A diversified and globally competitive ICT sector is born in Sweden

Not only Ericsson made layoffs in the turn of the millennium. In parallel a lot of .com companies and internet consultancy companies, which had been valued to fantasy figures during the hype, were in deep trouble. They lay off people or went into bankruptcy. The new millennium started with a complete turmoil, but out of the ashes raised a dynamic new ICT sector.

When Ericsson no longer absorbed most of the ICT competence in the country people had to turn somewhere else. Ericsson’s former employees looked for job elsewhere, but thousands of ICT consultants also had to look for new customers. That development actually started on a massive scale already in the mid-1990s when Ellemtel decided to close down the AXE-N project, one of the largest projects in Swedish industry so far. A lot of competent engineers working with R&D related software development turned to other employers or customers in different sectors, for example the engineering industry and the banking industry. Information and Communication Technologies are of great importance for industrial transformations (see Long in this volume).

Other more application oriented developers and entrepreneurs benefitted from the fact that Sweden and particularly Stockholm, much thanks to Telia and Ericsson, had a well-functioning infrastructure and citizens who were early adopters of new internet services. A lot of very successful new ventures, such as Skype, Spotify, MySQL, Unibet and Klarna, have been born in the segment Software and Net Services. Some of them are spun off from universities like KTH, but many are also spin offs from industry. In some sub-segments, such as the gaming industry, new ventures, like King, Mojang, Jice and Stardoll, seems to have a strong foothold among students. Most new ventures in the segment Software and Net Services seem to have one thing in common. They have not been reliant on contacts with Swedish incumbents or Swedish agencies (Giertz 2015:2).

When the bubble busted, in March 2000, it seems as if the explosion spread a lot of seeds which found good soil for new businesses and new ventures. But at the same time it changed the conditions for governmental efforts to support demand-driven research and new innovations in Swedish industry. Before 2000, incumbents like Ericsson, Televerket and ABB could set the agenda for both higher educations in engineering sciences and applied research in technical universities as well as in sectorial research institutes. But when companies like ABB and Ericsson decided to concentrate their business and close down their factories the existing research in fields like microelectronics and optics were left without their former demanding “customers”, who set the agenda for demand-driven applied research. A sectorial research institute like Acreo than became more of an incubator than an institute partly financed by Swedish incumbents with a common research interest. As a consequence,
governmental funding was also partly transferred to help the establishment of new ventures, mainly in the hardware sector.

**EU influences innovation policy and regional growth policy**

When Ingvar Carlsson announced his retirement in 1996 Göran Persson became the new Prime Minister. One important policy issue was to lower the unemployment rate. By then research and technology-based new ventures were regarded as important potential growth engines in Swedish policy. Around the turn of the millennium scientists and politicians pointed out that in Sweden we invested more in R&D than most other countries. Sweden was also ranking in top in terms of new patents. Yet we lay in the bottom of the OECD list when it came to new ventures and entrepreneurship. Many heavy commentators claimed it was due to an inability to commercialize research to reap the benefits of our outstanding research in the form of new enterprises in the country. As a consequence, great policy interest was concentrated to research – both in universities and in research institutes – in fields like ICT, biotech and energy. Seed capital, venture capital, incubators, patent counselling and the like were included in a new policy agenda.

The new agenda made the Social Democratic government form a new Swedish governmental innovation agency, VINNOVA (Verket för innovationssystem), on January 1, 2001. The new agency, which is sorting under the Ministry of Enterprise and Innovation, replaced part of the former governmental agency Nutek. VINNOVA got the assignment to administer state funding for research and development. The agency’s mission as defined by the government is to promote development of efficient and innovative Swedish systems within the areas of technology, transportation, communication and labour. The agency should accomplish this by giving financial aid to research-focused companies for research, development and legal costs. VINNOVA have contacts with and supports universities, research institutes and public sector organizations as well. VINNOVA also acts as a National contact agency for the EU framework program for research and innovation, which has in a way increased the funding.

You can argue that the public funding of research and innovation was increased when Sweden joined EU. An active policy to support the development in all regions in Sweden was even more boosted when Sweden joined the EU (Nilsson 2008). A broad political consensus came to prevail that Sweden should regain its share of the money paid to the common funds. Politicians and researchers, and representatives of business and trade unions, both at national and regional level, joined forces to make sure Sweden would be able to gain a rightful share of the common cake. Sometimes the discussions on development started rather in the availability of EU funds, than in discussions about actual needs and possibilities in the region. On the national level Nutek initially played a central role in helping the regions in this
context. But when Nutek was closed down in April 2009, the responsibilities were transferred to a new government agency, Tillväxtverket (The Swedish Agency for Economic and Regional Growth).

The Swedish Agency for Economic and Regional Growth is sorting under the Ministry of Enterprise and Innovation. The main task is to distribute EU funding to promote entrepreneurship and regional growth. The mission is to strengthen the competitiveness of Swedish small and medium-sized enterprises or future entrepreneurs directly and work to improve the general framework for doing business. The agency also has the ambition to build networks for cooperation and investment initiatives that hopefully will strengthen the business sector in the region.

Ambitions on the regional level have to some extent removed national policy focus from productivity issues. Former right-wingers among the Social Democrats, such as Kjell-Olof Feldt, had left. They made room for other ideas. A similar reversal was also made within the liberal-conservative block, which was to some extent manifested already when they regained power in 1991. Within both the two political blocks the key policy areas have lately to some extent been formed on the basis of other considerations than economic growth on the national level. The taxation emphasizes justice not incentives, labour law stresses security not mobility, and regional policy stresses equality in all parts of the country and so on. Growth policy does no longer permeate all policies on the national level. Instead growth policy has been looked upon as a separate policy area. Economic growth is no longer regarded a concern of general policy but the concern of a separate innovation policy and a regional growth policy.

Spiders – commercial network-builders in modern industry

The industrial dynamics in modern industry differs from the dynamics a century ago. In the old days Henry Ford was the role model for the recipe for success; to merge companies, to acquire suppliers and distributors in order to cover the entire chain from beginning to end in one single corporation. Today many companies concentrate on their core business and choose to procure services that they previously ran in-house. This leads to a more diversified business world and to more specialized companies, which must interact with one another in different business systems. The interaction is organized by spiders, who build different kinds of networks between more or less autonomous actors in the modern business world. While the spiders themselves account for an insignificant fraction of total employment, they are becoming increasingly important for large parts of modern business life. I differentiate between spiders of three different kinds; contractors, replicators and brokers (Giertz 2000).
When companies concentrate on their core business they partly turn from being employers to being contractors. As contractors they engage different service companies, contract manufactures, consultants, freelancers, suppliers and subcontractors to produce specific tasks. Contractors bring together various specialists in temporary projects. Typical contractors are project managers, turnkey construction contractors, book publishers, festival organizers and film producers. Contractors engage many people but have very few people employed. Researchers in Sweden started to observe contractors during the 1990s and called them imaginary organizations (Hedberg 2000). Today Contractors with an extreme business concept, who don’t have an office and are interacting with subcontractors mainly through telecommuting, are often called virtual organizations. Contracting is spreading as a consequence of outsourcing and spin-offs as well as increased specialization and professionalization of services. Relations between contractors and suppliers are often stable and long-term, even though each supplier is normally part of many different networks.

Outsourcing of different services of course also leads to the formation of new service companies and new service industries in the B2B sector. Many of these services, like machine repairs, installations and janitorial services must, just as many B2C-services, like car repairs, hair dressing or body care, be produced and delivered locally to the customer. But since companies and people in different locations have similar needs, it is possible to replicate one successful business concept from one location to another – and this is exactly what the second kind of spiders, replicators, do. The replicator is the central node in a replicating organization.

Replicating organizations, which are replicating a business model to several geographical locations, have during the last decades become a dominant type of organization in many different industries. One of the origins of replicating organizations can be found in the retailing sector and today domestic, international or global replicating organizations are dominating the retailing sector in developed countries. Replicating organizations are also since long very common in the hotel and restaurant industry, with McDonald’s as the global role model, and also dominating part of the transportation industry, like Taxi, and rental industry, like car rental. Domestic and international replicating organizations are spreading very rapidly and becoming more common in a great variety of local B2C as well as B2B services. This means that different local entrepreneurs more or less simultaneously can exploit different geographic markets with the same standardized business concepts. The building blocks of a replicating organization include brands, corporate identity programs, logotypes, designs, manuals and marketing and sales promotion.

Replicating organizations can be organized in many different ways. In some cases, the replicator at the top – the central node in a replicating organization
– owns both the business concept and all local branches. The US retail chain Walmart, with more than two million employees, is such an in-house retail chain. In franchise organizations the replicator at the top is a franchiser that owns the business concept but gives local entrepreneurs, franchisees, the rights to exploit the concept on a certain local market. In federative organizations similar local companies voluntarily get together to set up a replicator as common central node. Together they can decide upon a common brand and set up common purchasing and marketing organizations. In recent years a forth kind of replicating organizations, using internet and sharing economy, have been set up. Thus organizations like Uber and Airbnb compete with other more traditional replicating organizations. This forth kind of replicating organizations work more like brokers than traditional replicating organizations.

Some Swedish and Nordic replicating organization have been very successful on the international and global market during the last decades. Some are found among retail chains, i.e. H&M and Ikea. Both companies have built up their own product range. On the supply side they have worked as contractors buying manufacturing from many different suppliers. On the sales side, on the other hand, both companies own almost all their stores in-house in their global replicating organizations. Other successful replicating organizations are delivering local manual services on the global market. They started on the Nordic Market and expanded through acquisitions when the outsourcing trend went global. One example is the ISS, which offers facility management services and cleaning. In 2010 the group had more than 520,000 employees in more than 50 countries. Another example is the Swedish security group Securitas, which had over 280,000 employees in 45 countries in 2010.

The third kind of spiders, brokers, serves as intermediaries in transactions involving capital, goods or services, by matching a defined supply from sellers with a corresponding demand from potential buyers. Traditional brokers, such as insurance brokerage or real estate brokerage and telemarketing can be rather labour-intensive brokering activities and sales commissions usually account for the lion’s parts of revenues among those brokers. Other forms of traditional brokerage activities, such as freight brokerage, travel agencies and mail order firms are not as labour-intensive. They were traditionally based more on mechanical promotional efforts and passive order-taking. However, today traditional brokers in many different segments are outcompeted by new web-shops, web-portals and virtual stores. Internet-based brokers of various kinds increase the transparency and competition in many different industries. In Sweden Blocket, an online buying/selling website has completely changed brokerage on the Swedish second hand market. You select a location, item category, and then post your ad. It works well since almost everyone in Sweden uses it.
A Service society with many different business logics

The industrial dynamics in Sweden has over time dramatically changed where people work and what they do. Traditional categorizations of companies say essentially nothing about the unique features of different operations from the standpoint of organization, management, skill profiles or markets. This is especially true for different kinds of services – which becomes increasingly bad when the service sector completely dominates working life. Therefore, I have developed a catalogue of business logics which complement the traditional classifications into economic sectors or industries (Giertz 2000). My six main categories, which are broken down into 21 different business logics, focus on different organizational and management prerequisites.

It is obvious to anyone that management principles and success factors will vary from one business logic to another. There is a difference between organizing and managing the work of a steel mill or a manual assembly line, a tax office, a consulting firm, a retail store or a call centre. But despite these obvious differences many people do not specify what business logic they are referring to when they describe the evolution of management roles organizational development or leadership. In 2000 I, together with Staffan Larsson, analysed how employment in different business logics had developed over time in Sweden between 1970 and 1997 (Giertz & Larsson 2000). In 2011 I made a follow up until 2009.

At the beginning of the 20th century more than half the Swedish population made a living directly from what the soil, the forests, the mountains and the sea produced. Nowadays the conditions under which these fundamental livelihoods are pursued have changed dramatically. Agricultural yields have increased thanks to better farming methods and the work has been mechanized. Mechanization has also revolutionized other raw materials extraction. Thanks to combine harvesters, fishing trawlers, timber processors boomer rigs and other machines only 1.6 percent of the Swedish working population was needed to do all raw materials extraction in Sweden in the beginning of the 21st century.

Since the mid-1960s the manufacturing sector in Sweden has undergone a similar transition process as the sector of raw materials extraction. Mechanization, automation and restructuring have changed the need for labour – both quantitatively and qualitatively. The price of labour is highly important in industries that still employ numerous unskilled workers. As a result; many companies have located their labour-intensive manufacturing operations in low-wage countries. In parallel, mechanization and automation has reduced the work forces in remaining factories in Sweden as well. Put simply, semi-skilled repetitive tasks have been replaced by positions that require a broader understanding of operations, heavier responsibility and greater technical skills. The statistics clearly show the development. In 1970
still 29 percent of the Swedish working population was employed by a company producing some kind of goods. In 2009 no more than 13.8 percent of the work force in Sweden – both white collar and blue collar – were employed in companies producing any kind of goods, including everything from steel and pulp- and paper industries to engineering companies as well as breweries and bakeries.

Quite a lot of people are involved in the distribution of goods from factories to final customers. Originally, many factories were concentrated in coastal cities that had large ports for inbound and outbound shipments. Starting in the mid-19th century, the growing railway network gave rise to additional transport links. These were later supplemented by long-distance trucking and intensive air cargo services. Eventually the shipments led to the establishment of shops and department stores that could offer a growing variety of merchandise from all corners of the earth. Today, as the production of goods has become increasingly more concentrated, parts and components are also being shipped back and forth between factories in different parts of the world. The volume of goods shipments is thus constantly growing. Meanwhile many companies are trying to reduce their inventories and buffer stocks. They would prefer short cuts where goods are shipped directly from manufacturers to retailers – or even better, directly to the customers. Despite the heavy increase in transportations, the number of people employed in the transportation industries, e.g. trucking firms and sea lines, the transshipping industries, e.g. seaports, airports, terminals, warehouses and logistic companies, and the retailing industries, e.g. shops, stores, pharmacies and kiosks, has been slowly declining in Sweden, from 16.7 percent in 1970 to 13.8 percent in 2009.

In a modern society there are also basic common services that cannot be controlled by the wishes of individual users or their willingness to pay. Basic common services are shared resources. In almost every nation basic common services include public administration, the justice system, the national defence system, the police system, the customs service, fire-fighting services and the prison system. Basic common service also includes the operation of infra services of different kinds, such as roads, rail roads, water works, telecom operators and distribution systems for electricity. Basic common services also include organizations that safeguard common interests, e.g. the interests of all citizens, all municipal inhabitants or all members in for example a church, a confederation or a union. The share of the working population in Sweden that are employed by organizations delivering basic common services has been surprisingly stable over time, increasing from 8.1 percent in 1970 to 9.2 percent in 2009.

Besides distribution of goods and basic common services there is a very fast growing service sector. Services of various kinds – health care, child and elder care, consultancy work, restaurants, handicrafts, entertainment and so on – are providing gainful employment to more and more people in Sweden. The
share of Sweden’s working population engaged in delivering operational service production of various kinds increased from 36.2 percent in 1970 to 54.9 percent in 2009. One factor behind this increase is that today Swedish women work outside the home to as great an extent as men. Before WW II children and elderly were cared for by women working in the home but nowadays this work has become professionalized and salaried.

A considerable proportion of the service sector is directly aimed at us in our capacity as consumers, e.g. restaurants, hotels, day care centres, theaters, gyms, dentists, hairdressers, taxis, fast-food stands. The share of working people in Sweden that deliver direct services to consumers at local establishments increased from 12.8 percent in 1970 to 25.1 percent in 2009. Services aimed at companies and organizations (B2B) had roughly half as many employees in Sweden as consumer services (B2C) in the beginning of the 21st century but they were growing even faster. The reason for this was of course that companies and organizations increasingly choose to outsource professional services instead of supplying them in-house. One very fast-growing sub-sector was knowledge intensive services, e.g. consultancy firms and accountants, which increased its share of employment in Sweden from 2.9 percent in 1970 to 8.5 percent in 2009. Another large sub-sector was local manual services, such as various kinds of maintenance and repair work,
electrical installations, painting, plumbing, janitorial and security services. In 2009 11.3 percent of the working population in Sweden delivered local manual services of some kind to customers on the local market.

**Entrepreneurship is back on the policy agenda**

A few years into the 2000s a consensus grew on the desirability of a broad widespread entrepreneurship (Braunerhielm 2012). Not only an increased number of super entrepreneurs that would drive new technology-based companies. Entrepreneurship and business creation is an equally important ingredient to increase efficiency in different parts of working life – including domestic and local services. Fragmentation of organizations and companies into smaller units, and a subsequent streamlining of operations, could have significant impact on the pressure for change and growth. The emergence of a rich flora market funded, entrepreneurial enterprises can be of equal importance for economic growth as a single new company with high growth potential.

The wish to support new ventures and entrepreneurship made government introduce some changes in regulatory systems, for example in the tax systems. Swedish laws and regulations were in the beginning of 2000s still very much inspired by Harvard economist John Kenneth Galbraith, who advocated policies and regulations that favoured existing incumbents. He argued that individual efforts and individual incentives had become less important. That thinking was still reflected in the Swedish regulatory system in the beginning of the millennium. Something had to be done.

Sweden, compared to other countries, had a very high inheritance tax and gift tax, which were also taxed progressively. The inheritance tax and gift tax in Sweden caused many serious problems in connection with generation shifts in private companies, because private assets were often locked up in the companies. When assets were shifted, the heirs were often forced to either sell the company or take out a substantial private loan to cover the taxes. For several decades, the inheritance and gift taxes in Sweden, combined with a wealth tax – which also imposed taxes upon working capital in companies – forced many business owners to either sell their firm before transferring it to the next generation, or to move outside Sweden. Many financial advisors in Sweden had since the mid-1970s recommended successful Swedish business owners to take up residence in another country. All those taxes were repealed in the beginning of the new millennium. The inheritance and gift taxes were repealed on January 1, 2005 and the wealth tax on January 1, 2007. Taxes on residential property were also repealed in 2008. This not only panted the way for new entrepreneurs but also made successful Swedish entrepreneurs move back home to Sweden.
Managerial and policy implications

Globalization has during the last fifteen years fundamentally changed the rules and action opportunities for politicians. The financial market is global, both manufacturing and service companies operate in a global market. The companies that supply services to the infrastructure are international. Questions on corporate localizations of the development, production and back-office businesses are constantly up for discussion, as well as the exchange of labour in an increasingly open economy. Globalization has also opened for efficiency improvements and structural changes in almost all industries.

The industrial dynamics in manufacturing industries partly can be conceived as conflicting with an increased local entrepreneurship. Economies of scale are utilized and small companies are acquired and merged when industries are consolidating. In some mature manufacturing industries there are only a few companies left, which compete with one another in the global market. Alongside companies that expand on the global market are becoming more specialized.

Of course there are also new ventures born in manufacturing industries. Some of them origin from research and new technology and they might very well grow into successful global companies of importance for the export and trade balance in Sweden. But today only about 13 % of the people working in Sweden are employed by manufacturing industry – and the percentage is getting lower every year.

The Swedish work force has already moved from manufacturing industry to service industries of many different kinds. Thus a growth policy of tomorrow must not only – or even primarily – be about individual innovations or export companies with a high growth potential. Equally important – perhaps – is a policy which improves flexibility, efficiency and productivity in different service sectors that today accounts for 85 percent of all jobs in Sweden. Better quality and service with the same resources in all different service sectors is almost certainly of vital importance to grow the economy in Sweden. From a societal perspective, the main measure in this context might not be to get existing enterprises to employ more people. Perhaps it is even better to support the establishment of more market-funded enterprises that continue to finance their operations through revenues from services delivered on a competitive market.

One important question is if the business world’s ancillary systems in Sweden have kept up with the changes in industry. Are the parties on the labour market, authorities, ministries, legislative systems, insurance schemes, welfare systems, working hours, wage formation and so on adjusted to a
global, flexible and dynamic business life with different contracts and incentives in different industries? Or is a great part of the ancillary systems that were cemented in the 1970s – just when the steady growth of Swedish economy begun to turn back down – still controlling the agenda when government, authorities, employer associations, unions and others discuss future policy formulations in Sweden?

Another interesting question is if working life research and management theory in Swedish universities is adjusting to actual trends in today’s industry. Traditional management theory was formulated in close contact with manufacturing industry, which nowadays constitutes a very small part of working life. Some focus is definitely devoted to large incumbents in the engineering industry who integrate forward and become more of service organizations. Some focus is also devoted to services in the public sector, i.e. health care, and to the retailing industry. But the question remains: Is enough attention devoted to new management trends in different service sectors? Where can we in Sweden find courses on how to manage replicating organizations of various kinds or how to set up different net services on a domestic or global market or how to organize and manage successful knowledge intensive consultancy companies?

On methodological considerations and the use of sources in this chapter

All content in this chapter has been published before, mostly in Swedish, in various thesis, reports, books and articles. The main source is an anthology in Swedish that I edited in 2008. The anthology (Då förändras Sverige) contains 37 articles, with individual references, written by 25 Swedish researchers. Other important sources are reports in Swedish, written by myself and various co-authors, and published after assignments for different public authorities, ministries, industries, employer confederations, trade unions and research funding organizations in Sweden during more than 40 years. A few of these reports are mentioned in the reference list below. Thus this chapter is to a very large extent a very short conclusion and translation of my previous work and previous publication.

Suggested further readings

- For those who understand Swedish and want to get more into detail I strongly recommend to read the anthology Då förändras Sverige (Giertz (Ed) 2008).
• A classic book of great importance on industrial development in Sweden during two centuries is written by Lennart Schön: *Sweden’s road to modernity: An economic history.* (Schön 2011).
• For those interested in my catalogue of business logics I recommend *Measuring Success.* (Giertz 2000).
• Later development – especially in the ICT-sector – is presented in the anthology *Small and beautiful – The ICT success of Finland and Sweden.* Vinnova VA 2015:06.

**References**


Fridlund, M (1999) *Den gemensamma utvecklingen: Staten, storföretaget och samarbetet kring den svenska elkrafttekniken* (Stockholm)


Author presentations

Niklas Arvidsson

Niklas Arvidsson is Associate Professor in Industrial Dynamics at the School of Industrial Engineering and Management at the Royal Institute of Technology in Stockholm, and holds a Ph.D. in International Business from the Stockholm School of Economics. His research area focuses on innovation and learning within organizations and industrial systems with a particular focus on innovation in the area of payments. This is manifested in research projects studying the role of merchants in relation to payment services and omni-channel strategies, disruptive innovation in the payment industry, mobile payment services, the cashless society, and turbulence in the field of payments.

Pär Blomkvist

Pär Blomkvist, Ph.D., Historian of Technology and Docent at INDEK. He is doing research in the fields of industrial dynamics and management of innovation in large sociotechnical systems, with a special interest in evolutionary theories on technical change and institutional factors influencing the process of system dynamics. He is involved in research projects on investments in the energy sector (Energimyndigheten), on water systems in African cities (FORMAS) and air traffic systems transformation in Sweden (Trafikverket).

Eric Giertz

Eric Giertz, Ph.D., is a Professor emeritus in Industrial Economics and Management at KTH. He is specialized in Business Development linked to structural changes in industry. For 20 years Giertz worked as a strategy consultant in the field of business development. His clients included companies in manufacturing and service industries as well as ministries and authorities. He has experience as member of executive boards in companies in various industries and as chairman and part owner of successful new ventures. He is also a member of the IVA, the Royal Swedish Academy of Engineering Sciences, and the former Chairman of its Management Division.
Rurik Holmberg

Rurik Holmberg holds a Ph.D. from programme Energy Systems at Linköping University. His thesis focused on the concept of path dependence in the non-conventional oil sector. He graduated from Åbo Akademi University (Finland) in economics. Currently he is working as an analyst at the Swedish Energy Agency, with a focus on policies for energy efficiency in Sweden and internationally.

Petter Johansson

Petter Johansson is a researcher with focus on value network dynamics, firm strategy and systems theory in relation to socio-technical transitions in the energy sector. His empirical studies are focused on the industrial and societal transformation in the Swedish heating sector from the 1970s and onwards and the transition from oil dependence to renewable energy sources in the Swedish energy system, with a special focus on the heat pump technology. He teaches in energy related and industrial dynamics oriented courses at KTH.

Emrah Karakaya

Emrah Karakaya is postdoctoral researcher and lecturer at the unit of Sustainability and Industrial Dynamics at INDEK. He holds a double Ph.D. degree in Industrial Management (from Technical University of Madrid & KTH Royal Institute of Technology, 2015). His academic and empirical expertise includes industrial dynamics, sustainability, renewable and non-renewable natural resources, diffusion of innovations and business models. He has taught several classes at MSc level in the fields of Innovation Management and Energy Business.

Vladimir Kutcherov

Vladimir G. Kutcherov graduated from the Gubkin Russian State University of Oil and Gas (Moscow) as a petroleum engineer in 1977. He is currently a Researcher at the School of Industrial Engineering and Management at the Royal Institute of Technology (Stockholm), and a Professor at the Department of Physics at Gubkin Russian State University of Oil and Gas (Moscow). Mr. Kutcherov is a member of the Deep Carbon Observatory researcher community. His main research interests involve the fundamental problems of high-pressure organic chemistry, innovations in the energy sector, and efficiency and environmental aspects of energy transformation processes.
Staffan Laestadius

Staffan Laestadius is prof.em. of Industrial Dynamics at INDEK/KTH, and a member of the board and senior advisor at the think tank Global Utmaning. His research in industrial dynamics has been focussed on the mechanisms of industrial and technical change. This includes analyses of the knowledge base and non-science based innovations and mechanisms of change in “low-tech industries” like the forest based industry. During recent years Laestadius has analysed the conditions for the Swedish and European economies to combine industrial and social transformation with climate change mitigation and adaption and still maintain the “Nordic model of welfare”.

Vicky Long

Vicky Long, Ph.D., is a researcher at the Department of Industrial Economics and Management (INDEK) at the Royal Institute of Technology (KTH), Sweden. Her research area is technology, innovation and development. She has studied, at both a meso (industrial) level a micro (firm) level, the dynamics of technological learning and capabilities creation in a context of industrial upgrading and globalization. Empirically, she has extensively studied Information and Communication Technologies (ICTs) whereas Intellectual Property Rights (IPRs) play important role in knowledge sourcing and knowledge appropriation. Her new empirical explorations embrace industry sectors in renewable energies and transport.

Maria Morgunova

Maria Morgunova is currently a Ph.D. candidate at KTH Royal Institute of Technology at the Department of Industrial Economics and Management. In 2015 Ms. Morgunova received the degree of Licentiate at KTH with her thesis “Arctic Offshore Hydrocarbon Resource Development: Past, Present and Vision of the Future”. Her research interests include the economics and management of energy resources, energy markets transformation, the Arctic region and Arctic hydrocarbon resources development.

Michael Novotny

Michael Novotny recently defended his thesis in Industrial Economics and Management at the division of Sustainability and Industrial Dynamics (KTH Royal Institute of Technology). His thesis “Breaking the chains: A technological and industrial transformation beyond papermaking – Technology management of incumbents” focused predominantly on technology shifts and new business models of pulp and paper industries mainly located in the Northern hemisphere, that are trying to transform due to new market conditions and to introduce new bioproducts based on woody biomass.
Cali Nuur

Cali Nuur is a professor in Industrial Dynamics and the head of the Sustainability and Industrial Dynamics (SID) division at the Royal Institute of Technology, KTH, Sweden. He teaches and conducts research in industrial and technological transformation processes and the mechanisms that underlie them. Theoretically, Cali’s research is underpinned by the discussions on “the new production of knowledge” which has prevailed among academics in recent decades. A key theoretical point of departure for his research is innovation theory with emphasis entrepreneurship, learning, knowledge building processes, competitiveness, capability creations and diffusions etc as vital mechanisms of development.

Annika Rickne

Annika Rickne is Professor of Innovation, Entrepreneurship and Industrial Dynamics (guest) at the Department of Industrial economics and management, at the Royal Institute of Technology (KTH) in Stockholm. She earned her Ph.D. in Technology Management from Chalmers University of Technology in 2000, and was named reader (Docent) at Lund University in 2005. Her research interests include the dynamics of the ICT sector, policies and strategies for regenerative medicine in California, customer involvement in innovation, and how and why academic scientist in Turkish nanotechnology engage in industrial activities.

Pranpreya Sriwannawit Lundberg

Pranpreya Sriwannawit Lundberg is a policy developer at National Science Technology and Innovation Policy Office in Bangkok, Thailand. She is also a ministerial staff serving under Thailand’s Minister of Science and Technology, focusing on international affairs. Pranpreya holds a Ph.D. in Industrial Economics and Management from KTH Royal Institute of Technology in Stockholm, Sweden. Her doctoral dissertation is titled “Power to the People: Diffusion of renewable electricity in rural areas of developing countries”.

372