Technological Transitions and System Innovations

Technological Transitions and System Innovations

A Co-Evolutionary and Socio-Technical Analysis

Frank W. Geels

Assistant Professor, Department of Technology Management, Eindhoven University of Technology, The Netherlands

Edward Elgar

Cheltenham, UK • Northampton, MA, USA

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

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

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

ISBN 1 84542 009 8

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

Contents

Foreword

Anyone who sets out to understand the dynamics of long-term technological transitions faces a huge challenge. The range of material is immense and varied. Economic and sociological theories of technical change need to be incorporated, although these disciplines do not have a strong tradition of relating to each other. Responding to a challenge like this could thus easily lead to a flight into the realms of concepts or the world of empirical descriptions. Genuine understanding will be substituted for a very high level of generalisation, with reference to superficial empirical work, or the reverse will happen: a lot of data will be produced without showing what they imply, using the discipline of theory.

Frank Geels has succeeded in avoiding both traps. He has brought together in a meaningful way a rich account of several cases in the context of an attempt to develop a new theoretical perspective. His analysis deals with a substantial issue: the role of technical change in our modern society. This issue arouses passion, and rightly so, because technical change is an important but contested driving force. Advancing our understanding of such a force is one of the best services that scholars of innovation can offer to society. This book introduces a balanced, comparative and critical survey of a range of different perspectives on long-term technological change. This represents a substantial resource for any reader who seeks to become informed on recent theories in the field. However, it goes beyond presenting current work, and the book can be seen as an innovative intervention in the field. Hence, it will interest the reader who wants to know about a new and promising way of looking at technological change. It deconstructs some of the received and classical accounts of major transitions, for example of horses by gasoline cars, and sailing ships by steamships. Geels complicates the often too simple stories of substitution, and leads us to a new understanding of path dependencies through the introduction of a multi-level perspective which incorporates broader change processes and a rich analysis of the historical context in which new technology emerges. He shows that the search for larger patterns in a complex transformation process can be rewarding. His book is an important contribution to the field of innovation studies, including the sociology of technology, evolutionary and institutional theories of technical change in economics and the history of technology. It is an invitation to make the next step: from doing individual case studies and developing piecemeal insights to seeking a deeper understanding of patterns in the direction and processes of technological transitions that span 50 to 100 years.

Johan Schot
Professor in the History of Technology
Department of Technology Management
Eindhoven University of Technology
The Netherlands

Preface

This book is about technological transitions and system innovations. These processes refer to changes in the way societal functions are fulfilled, for example, transport, communication, housing and energy supply. Societal functions are fulfilled by socio-technical systems, consisting of technology, markets and user practices, public policies and regulations, infrastructure, symbolic meaning and scientific understanding. The change from one socio-technical system to another is the topic of this book. Because technology is a crucial element in socio-technical systems, it is taken as an entry point to study system innovations. But change is not limited to technological transitions. There are all kinds of knock-on effects, which affect the entire socio-technical system. Furthermore, I conceptualise technology not just as knowledge and artefacts. Working from the discipline of science and technology studies, I see technology as heterogeneous configuration of elements that are aligned and work together. Social and technical aspects are always intertwined and constitute each other. This book analyses how social and technical aspects influence each other during technological transitions and system innovations. Hence, the subtitle of the book: a co-evolutionary and socio-technical analysis.

The book develops a conceptual perspective to understand how system innovations come about through the alignment of social and technical developments at multiple levels. This qualitative perspective integrates insights from several disciplines, for example, evolutionary economics, innovation studies, sociology of technology and complex systems theory. The conceptual perspective is multidisciplinary, even though science and technology studies form the basic conceptual background into which insights from other disciplines are integrated. The book also proposes more fine-grained socio-technical patterns and mechanisms and tests them with three historical case studies from the transport domain.

This book is an adapted version of my PhD thesis, which was called 'Understanding the dynamics of technological transitions'. Some readers of the thesis commented that both the analysis and the case studies were wider than the title suggested, and rather looked at changes in socio-technical systems. I have taken this comment into account when rewriting the text for this book. I have not only changed the title, but also included system approaches in the literature review, something which offered additional building blocks for the conceptual perspective. Furthermore, I have substantially shortened the original text and created a better structure and storyline. Since completing my thesis, I have been able to take a step back from the text. With the benefit of hindsight I had few problems with leaving out side issues, thereby improving the overall flow.

I would like to thank several people for providing a stimulating context for writing this book. Johan Schot and Arie Rip have been essential in their role as supervisors of my PhD thesis: we have had stimulating discussions and debates; they acted as sparring partners for the development of ideas and concepts in this book; and their intuitions about fruitful research directions were indispensable to open up new vistas when I got stuck. I would like to thank René Kemp for ongoing discussions and collaborative work. My former colleagues at the Department of Philosophy of Science and Technology (FWT) at the University of Twente have provided a stimulating and intellectually diverse context. In particular I want to thank Boelie Elzen and Peter Hofman for pleasant work in shared projects, and Ken Green from Manchester University for his stimulating comments during his visits. Together with my new colleagues at the Department of Technology Management at the University of Eindhoven we are working on a research program about transitions and system innovations. The theoretical perspective developed in this book is one of the guiding principles in this research. We use this perspective to study both ongoing and future transitions, as well as historical transitions. In particular, we have set up a 3-research project for several postdocs to create a database of about 30 case studies of historical transitions and system innovations. We will report about these case studies in the following years, and use them to further refine the theoretical perspective developed in this book. Geert Verbong, Rob Raven, Johanna Ulmanen, Johan Schot, Deborah Tappi and myself have the ambition to become a high-quality transition research group. As such, we will contribute to the newly created Knowledge Network on System Innovation (KSI) in the Netherlands. KSI is funded by a major investment of the Dutch government (10 million euro), which sees system innovations and transitions as a strategic research topic with wide societal relevance. We look forward to stimulating interactions, discussions and collaborations in the KSI program, set up and managed by Jan Rotmans, Johan Schot and John Grin. The availability of financial resources and inspiring researchers means that research into transitions and system innovations has a bright future. This book aims to contribute to that future.

I have tried to obtain permissions for the use of copyright material. For some of the older figures this was difficult, because some publishers no longer exist or have merged with other publishers. If people or organisations have been inadvertenly overlooked, I sincerely regret this and request them to contact me.

Eindhoven
September 2004

1. Introduction

1.1 THE TOPIC OF ANALYSIS

System Innovation

This book is about transitions from one socio-technical system to another. It is about system innovation, which includes but is wider than product and process innovation. Socio-technical systems can exist at different levels. In organisation and management theory the focus is at the level of firms. The interdependence of the social and technical systems of organisation was one of the core insights of the socio-technical systems tradition associated with the Tavistock School (Trist and Murray, 1990, 1993; Emery, 1993; Trist et al., 1997; Griffith and Dougherty, 2001). This tradition highlights that human beings and machines are joined together in industrial workplace settings, for example, coalmines (Trist and Bamforth, 1951). In this book, however, the focus is on a different level, namely, that of societal functions such as transportation, communication, housing, energy supply, recreation and health care. Technologies and artefacts play an important role in fulfilling these societal functions. Ships, cars and aeroplanes are necessary for transportation; telephones and the Internet make long-distance communication possible. But technologies do not fulfil societal functions on their own. Artefacts by themselves have no power; they do nothing. Only in association with human agency, social structures and organisations do artefacts fulfil functions. In real-life situations (for example, organisations, firms, houses) we never encounter artefacts *per se*, but rather artefacts in context (Fleck, 1993, 2000). For the analysis of functioning artefacts in context, the combination of 'the social' and 'the technical' is the appropriate unit of analysis. At the level of societal functions, a range of elements are linked together to achieve functionality, for example, technology, regulation, user practices and markets, cultural meaning, infrastructure, maintenance networks and production systems. This cluster of elements is called a 'socio-technical system', thus highlighting that social and technical aspects are strongly interlinked. Figure 1.1 gives an example for the socio-technical system for road transport.

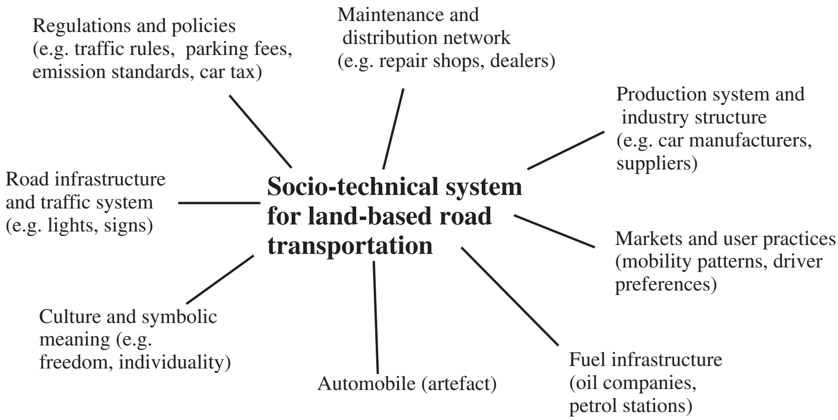


Figure 1.1 Socio-technical system for road transport

Examples of System Innovation

A system innovation is a transition from one socio-technical system to another. An example is the shift from horse-drawn carriages to automobiles in land transport. This transition obviously involved changes in artefacts, but also in infrastructures, for example, the creation of a gasoline fuel infrastructure and a road network which could carry heavy cars and trucks. It also involved cultural changes, for example, an appreciation of speed, excitement (touring in the countryside) and freedom (to determine one's own route in contrast to fixed tram and omnibus lines). Regulation was involved, because the street became more chaotic as all kinds of vehicles intermingled (bicycles, cars, trams, horse-drawn wagons), leading to accidents. Traffic rules were devised to create order (for example, rights of way, speed limits, driving licences). Policy makers had to make decisions about the allocation of funds for road building, at the expense of investments in an electric tram infrastructure. The rise of automobiles led to a new industry with many forward and backward linkages in the economy. The automobile industry pioneered new methods of mass production (assembly lines, Fordism). All these elements co-evolved and were intermingled, as will be described in Chapter 5.

Another example is the transition from sailing ships to steamships. This involved a change from wood to iron and from sail to steam engines and screw propulsion. Steamships increased speed and carrying capacity, but were more expensive to build and had high coal costs. Steamships also introduced new specific functionalities in shipping: regularity and predictability. This was possible because steamships were independent of winds. This facilitated a new operating practice, liner companies, operating on the basis of fixed departure and arrival schedules. Liners were operated in fleets, which required new management skills (fleet management). To raise money for the fleets, a new institution was developed, the joint-stock company, separating management and ownership. To make their functioning possible, a worldwide coal infrastructure was created so that ships could refuel in ports. As ships became ever larger, ports were enlarged and deepened. Because of their high capital cost, ships had to be unloaded quickly. Hence, new cargo-handling machines were introduced on docks (for example, cranes, elevators). Here, too, many elements co-evolved and intermingled.

Technology as Entry Point

Although many elements are involved in system innovations, it is easy to name them on the basis of the main technologies involved, for example, the transition from horse-drawn carriages to automobiles or from sailing ships to steamships. This is partly a matter of naming, and partly a matter of the tendency to look at tangible artefacts first. But I argue that it is also because technology plays an important, although not determining, role in system innovations: technologies and artefacts have assumed an incessant endogenous innovative dynamic in modern, capitalist societies. Both the disciplines of various technologies as well as their materialisations in the form of artefacts have become the object of ceaseless and organised change. Technical innovation has become institutionalised in research and development (R&D) laboratories, universities, technical institutes, patent laws and so on. Furthermore, technical innovation is an important battleground where firms compete. This emergent web of technical novelty may well be a destabilising force for socio-technical systems. These considerations do not mean that technology is always the prime mover of system innovations. I am not a technological determinist. But it does mean that technology is very likely to play a prominent role in many system innovations. Technology is a crucial part of the transition, not an external input, as it is often regarded. Hence I argue that technology is a promising entry point to study system innovations. This is a sensible choice in the light of research strategies. Because socio-technical systems are complex, it is wise to limit one's ambitions. I prefer to study socio-technical systems from a particular angle, rather than in their entirety. Technology is a good entry point as I argued above. This also means that I can build upon the large literature on technological change, using it as a platform to explore system innovations. But I shall not limit myself to looking only at products, artefacts, production technologies, product substitutions and so on. I aim to analyse how technical changes interrelate with other elements in system innovations. This aim is indicated by the title of the book, *Technological Transitions and System Innovations*, which highlights that the shift from one technology to another is an important part of system innovations (for example, from sailing ships to steamships).

Position with regard to Innovation Typologies

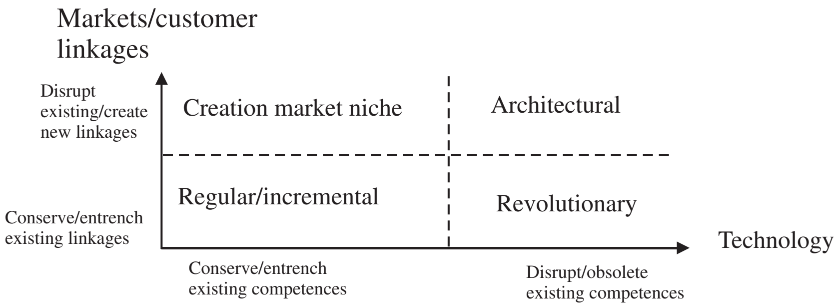
To further delineate the topic of analysis, it will be positioned with regard to two innovation typologies, one by Freeman and Perez (1988) and the other by Abernathy and Clark (1985).

Freeman and Perez distinguish four kinds of innovations. 'Incremental innovations' occur more or less continuously in any industry or service activity depending upon a combination of demand pressures, socio-cultural factors, technological opportunities and trajectories. They are very important for improving the efficiency of all factors of production. 'Radical innovations' are discontinuous events which are unevenly distributed over sectors and over time. Whenever they occur they are important as the potential springboard for the growth of new markets, and for the surges of new investment associated with booms. They often involve a combined product, process and organisational innovation. Over a period of decades radical innovations, such as nylon and 'the pill', may have fairly dramatic effects in bringing about structural change. In recent times they are usually the result of a deliberate R&D activity in enterprises and/or university and government laboratories. Changes of 'technology system', are far-reaching changes in technology, affecting several branches of the economy, as well as giving rise to entirely new sectors. They are based on a combination of radical and incremental innovations, together with organisational and managerial innovations affecting more than one or a few firms. Changes in the 'techno-economic paradigm' (TEP) are so far-reaching in their effects that they have a major influence on the behaviour of the entire economy. A change of this kind carries with it many clusters of radical and incremental innovations, and may eventually embody a number of new technology systems. The changes involved go beyond engineering trajectories for specific product or process technologies and affect the input cost structure and the conditions of production and distribution throughout the system. Some examples of TEP are steam power and iron (1830–80); electricity and heavy engineering with steel (1880–1930); automobiles, aircraft, oil and petrochemicals and synthetic materials (1930–80); and computers, software and telecommunications (1980–?). With regard to this typology, system innovations and technological transitions are close to changes in 'technology systems'. They differ from changes in TEP because the latter refer to pervasive technologies on the level of entire societies, while system innovations occur on the level of societal functions.

Abernathy and Clark develop their typology by distinguishing two dimensions. The first consists of linkages (of a firm) with customers. Elements on the market/user side involve: relationship with the customer base, customer applications, channels of distribution and service, customer knowledge and modes of customer communication. An innovation can disrupt existing linkages and create new ones, or conserve/entrench existing linkages. The second dimension relates to the technology and production competences (of a firm), which can be disrupted and made obsolete or conserved and entrenched by an innovation. Elements on the technology/production side involve: design/embodiment of technology, production systems/organisations, skills (labour, managerial, technical), materials/supplier relations, capital equipment, knowledge and experience base. Crossing these two dimensions results in four types of innovation (see Figure 1.2):

1. *architectural*: disrupts technology/competence *and* linkages with users;
2. *creation market niche*: conserves technology/competence but breaks linkages with users;
3. *regular/incremental*: conserves both technology/competence and user linkages; and
4. *revolutionary*: disrupts technology/competence but conserves user linkages.

In the definition of Abernathy and Clark (1985), architectural innovations (AIs) involve both changes in the technology and changes on the user side, introducing new functionalities and user practices. Abernathy and



Source: Abernathy and Clark (1985: 8); courtesy of Elsevier Science.

Figure 1.2 Typology of innovations

Clark developed their innovation typology primarily to determine the consequences of different kinds of innovation for firms. While their point about linkages and alignments between elements is important, their focus on firms is too limited for my research topic. Disruptions occur on a much wider scale during system innovations. The functional side of socio-technical systems not only involves users and markets (as in Abernathy and Clark's definition), but also public policies, symbolic meanings, infrastructures, maintenance networks and so on. Thus, system innovations can be described as AIs writ large. System innovations involve changes on the technical side *and* the user side. Without changes on the user side, technological discontinuities are better described as 'technological revolutions', which change the technical dimension of socio-technical systems, but leave the rest relatively unchanged.

1.2 RESEARCH QUESTIONS, ACADEMIC AND SOCIETAL RELEVANCE

The following three research questions are addressed in this book:

1. How do system innovations and technological transitions come about?
2. Are there patterns in system innovations and technological transitions?
3. Are there particular mechanisms in system innovations and technological transitions?

I make the following distinction between patterns and mechanisms: while patterns are outcomes, mechanisms produce outcomes. Mechanisms can play a role speeding up/slowing down processes, or lead to changes in direction. Furthermore, patterns typically stretch over the entire process of system innovation, while mechanisms take place over shorter time periods.

Academic Relevance

With the focus on patterns and mechanisms, I can make a new contribution to science and technology studies (STS). In STS there has been much attention on contingency, complexity and heterogeneity, often using a micro focus on actions and interactions between individual actors in local practices; little heed has been paid to long-term processes and patterns.[1] Following the actors is important, but one is then reduced to embracing complexity and contingency. However, local interactions add up to patterns on a more aggregated level. Identifying such processes and patterns in the complexity of system innovations is one research aim.

System innovations are a complex topic, which has not yet received much attention. In the STS discipline there is a research school that looks at large technical systems (LTS), such as electricity networks, railroad networks, telephone systems, videotex and the Internet (Hughes, 1983; Mayntz and Hughes, 1988; La Porte, 1991; Coutard, 1999). Hitherto, LTS research has focused mainly on the emergence and development of large technical systems, rather than on the change from one system to another, although Summerton (1994) made a start. Other studies in STS looked mainly at the development of single technologies and artefacts, for example, bicycles, bakelite and bulbs (Bijker, 1995) and aircraft projects (Law and Callon, 1992). To understand these developments, sociological and socio-technical analyses have been made, but socio-technical systems as such have not been studied.

In innovation studies, most work has focused on firms and industries (for example, R&D, knowledge management, learning in networks and firm strategies), and not on socio-technical systems. Incremental and radical innovations have been examined, often with regard to their implications for technology management (for example, Tushman and Anderson, 1986; Anderson and Tushman, 1990; Christensen, 1997), but system innovations have barely been addressed, even though Freeman and Perez (1988) saw it as one of four kinds of innovation. There is a new research line in innovation studies, however, which takes up the challenge to some extent: the systems of innovation approach. In the literature we find national systems of innovations, regional systems of innovation, sectoral innovations systems and technological systems (see discussion in Chapter 2, Section 2.2). The systems of innovation approach looks at technical innovation, and argues that it emerges not just from individual firms, but from a wide network or system of actors (universities, research institutes, government programmes, R&D departments in firms). An important insight from the systems of innovation approach is the emphasis on interlinkages between elements and the co-evolutionary process. The main focus in this approach is on the *functioning* of systems, for example, static or comparative analyses of the innovative performance of different countries and sectors, but the *change* of systems has largely been ignored. System innovation is thus an interesting research topic.

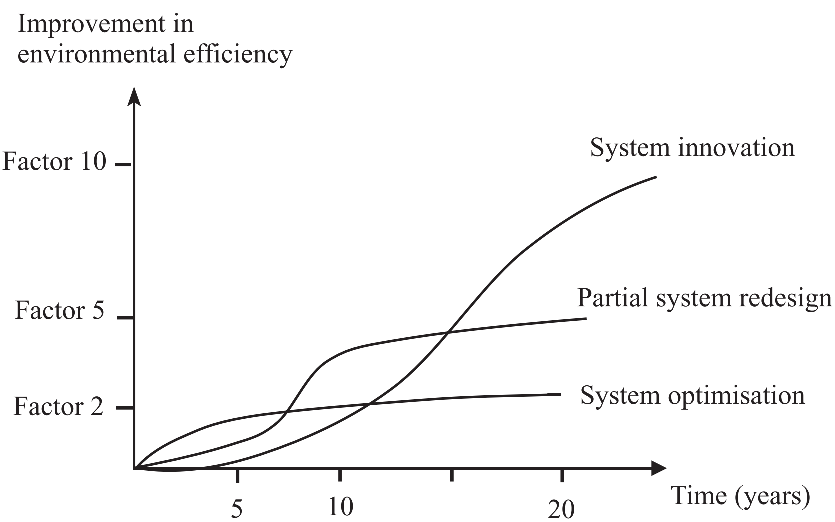
The research questions in this book are also relevant with regard to path-dependence theories and evolutionary economics. David (1985) and Arthur (1988) have shown that path dependence plays an important role in the case of two competing technologies. Once one of the technologies has gained a lead, it benefits from increasing returns to adoption and creates a dominant path. Several mechanisms cause increasing returns, such as economies of scale leading to lower cost, learning by using, network externalities, informational increasing returns and technological interrelatedness. Because of these increasing returns a certain technology becomes entrenched. Other economists have widened the analysis, adding institutional aspects and user routines to the lock-in analysis (Cowan, 1990; Cowan and Gunby, 1996). Other disciplines have added more reasons why existing systems are characterised by stability, inertia and lock-in. Established systems may be stabilised by legally binding contracts (Walker, 2000). Actors and organisations are embedded in interdependent networks (with suppliers or users), which represent a kind of 'organisational capital', and create stability through mutual role expectations. Evolutionary economists argued that cognitive routines make engineers and designers look in particular directions and not in others (Dosi, 1982; Nelson and Winter, 1982). This can make them 'blind' to developments outside their focus. Core capabilities can thus turn into core rigidities (Leonard-Barton, 1995). Evolutionary economists coined concepts such as 'technical paradigms' (Dosi, 1982) and 'technological regimes' (Nelson and Winter, 1982) to explain the stability of established technological trajectories. STS researchers emphasise that incumbent systems are also stable because they are embedded in society. People adapt their lifestyles to them, favourable institutional arrangements and formal regulations are created, and accompanying infrastructures are set up. The alignment between these heterogeneous elements leads to technological momentum (Hughes, 1994). The importance of these alignments between heterogeneous elements is highlighted in such concepts as the 'techno-institutional complex' (Unruh, 2000) and 'techno-economic networks' (Callon, 1991). All these approaches highlight aspects of stability and path dependence of existing systems but none of them addresses the issue of change and transition from one system to another. Given all these explanations of stability, it is a mystery how and why transitions occur. Path-dependency literature may help us understand lock-in, but how can we understand 'lock-out'? That is the topic of this book.

Societal Relevance

The topic of system innovation is not only academically interesting, but also has societal relevance. Modern societies face structural problems in several sectors. In the energy sector there are problems related to oil dependency, reliability, and CO_2 and NO_x emissions. The transport system suffers from congestion, air pollution (particulates, NO_x), energy use and CO_2 emissions. Cattle farming suffers from manure disposal problems, ammonia emissions and diseases like BSE (bovine spongiform encephalopathy) and foot and mouth. These problems are deeply rooted in social production and consumption patterns.

In the past two decades, much effort has been made to solve problems with product and process innovations. Cleaner products and processes have been developed alongside the application of end-of-pipe solutions. Sometimes these innovations have led to substantial improvements in environmental efficiency, such as in the case of automobile catalysts which greatly reduced tailpipe emissions of pollutants. Substantial improvements in environmental efficiency (a 'Factor 2' is a general average) may still be possible with innovations of an 'incremental' kind. But larger jumps in environmental efficiency (possibly by a 'Factor 10') may only be possible with system innovations, for example, transport systems based on hydrogen and fuel cells. The promise of jumps in environmental efficiency via system innovations is schematically represented in Figure 1.3.

Because of this sustainability promise, there is increasing interest from policy makers, non-governmental organisations (NGOs) and large firms in transitions and system innovations. For instance, the Stockholm Environment Institute has published *Great Transition* (Raskin et al., 2002); the US National Research Council (1999) and the Dutch Research Council (NWO) have made the study of transitions part of their research portfolio; the IHDP research programme (International Human Dimensions Programme on Global Environmental Change) has a project on 'Industrial Transformation' (similar in meaning to 'transitions'); and the



Source: Weterings et al. (1997).

Figure 1.3 Environmental efficiency and system innovation

Dutch government gave transitions a central place in its fourth National Environmental Policy Plan (VROM, 2001) and has established 'transition teams' within various ministries.

To link with and to feed this growing interest, this book explores how system innovations come about and how policy makers might influence them.

1.3 BASIC ASSUMPTIONS AND CONCEPTUALISATIONS

Because many heterogeneous elements change during a system innovation, it is a co-evolutionary process. These changes involve technical innovations *and* innovations on the user side. As promised in the subtitle, this book makes a co-evolutionary and socio-technical analysis. Co-evolution refers to the wider process of system innovation. Socio-technical refers to a particular analytic perspective, which works from several basic assumptions and conceptualisations of technology, human action and social structure. These assumptions and conceptualisations are made explicit in this section.

Conceptualisation of Technology

Because technology is chosen as the entry point to study system innovations, it is important to use a conceptualisation of technology that is appropriate for the task. This means that it should match with the topic of analysis and conceptualise socio-technical linkages. Such a conceptualisation can be found in the discipline of STS, which can stand for either science and technology studies or science, technology and society. The former emphasises the social conditioning of technical and scientific development as itself a social activity, whereas the latter is more concerned with the societal consequences of science and technology. STS is a heterogeneous discipline, which builds upon insights from the history and philosophy of science and technology, sociology of scientific knowledge, sociology of technology, political theory and economics of technical change. With regard to technological development STS consists of a range of schools, for example, social shaping of technology (SST), social construction of technology (SCOT), actor–network theory (ANT) and large technical systems theory (LTS). Although these schools have different emphases, they share two basic notions of technology: (a) technology is heterogeneous, not just a material contraption; engineers know this, their work is 'heterogeneous engineering' (Law, 1987); and (b) the functioning of technologies involves linkages between heterogeneous elements. Hughes (1987) coined the metaphor of a 'seamless web' to indicate how physical artefacts, organisations (for example, manufacturing firms, investment banks, R&D laboratories), natural resources, scientific elements (for example, books, articles) and legislative artefacts (for example, laws) are combined in order to achieve functionalities. In actor–network theory (for example, Latour, 1991, 1993; Law and Callon, 1992) these linkages between elements are taken as a basic ontology. There are no elements without linkages. The boundary between 'social' and 'technical' is created in articulation and translation processes. Latour (1991), for instance, writes:

> We are never faced with objects or social relations, we are *faced with chains* which are associations of human (H) and non-humans (NH). No one has ever seen a social relation by itself . . . nor a technical relation . . . Instead we are always faced by chains which look like this: H–NH–H–NH–NH–NH–H–H–H–H–NH . . . Of course, an H–H–H assembly looks like social relations while an NH–NH–NH portion looks like a mechanism or a machine, but the point is that they are *always integrated into longer chains*. It is the chain (the syntagm) we study or its transformation (the paradigm), but it is never some of its aggregates or lumps. (p. 110, added italics)

> You are faced with . . . shifting assemblies of associations and substitutions. . . . We observe a process of translation. (p. 116)

Rip and Kemp (1998) propose the pragmatic definition of technology as 'configurations that work'. While the term 'configurations' refers to the alignment between a heterogeneous set of elements, the addition 'that work' indicates that the configuration can and should stabilise in 'fulfilling a function'. This definition of technology is important, because it encompasses both the configuration and the functional aspects. Functions are fulfilled because elements of a configuration are aligned to one another. In the case of *new* technologies the conceptualisation refers to configurations that *should* work in the end, if enough effort is put into developing it. Technical configurations are shaped in a particular form *to fulfil a function*. This alignment between (technical) form and (social) function is indicated as the 'dual nature of technological artefacts' (Van de Poel, 1998: 12). More aspects than mere technical form are needed to realise the functionalities (for example, organisations, infrastructures, regulations). Thus, technologies are always 'technologies-in-context' (Rammert, 1997: 176). Artefacts are never used in a vacuum, but always in an application domain (for example, a household, a transportation system), which is structured, for example, by regulations, user practices, symbolic meanings and maintenance organisations. This context is not merely passive, but actually helps the artefact to fulfil a function.

This conceptualisation of technology as heterogeneous 'configurations that work' is useful to analyse system innovations, because it emphasises the inherent linkages between technical and social aspects. This conceptualisation can be used at different levels, ranging from firms to socio-technical systems at the level of societal functions. STS thus is a good candidate for making a socio-technical analysis of system innovations.

Human Actors and Social Structure

There are two fundamentally different conceptions of the activities of human actors. In the first, human actors are viewed as the essential sources and forces of social change. The individual, the strong personality as exemplified by Joseph Schumpeter's entrepreneur or Thomas Hughes's system builder, enjoys an extensive freedom to act. In the second conception, social actors are faceless automata following iron rules, institutions or given roles/functions in social structures which they cannot basically change. While the first view emphasises agency, the second highlights the effects of structures.

In recent decades, conceptual approaches have been developed which attempt to solve the structure–agency dilemma (for example, Bourdieu, 1977; Giddens, 1984; Burns and Flam, 1987). In these approaches actors are seen as embedded in wider structures, which configure their preferences, aims and strategies. Despite these structuring effects, the approaches leave much room to actors and agency, that is, conscious and strategic actions. Giddens, for instance, talks of the 'duality of structure', where structures are both the product and medium of action. Bourdieu coined terms such as 'habitus' and 'field' to conceptualise similar notions. And Burns and Flam developed a 'social rule system theory' to understand dynamic relationships between actors and structure. In all these approaches human agency, strategic behaviour and social struggles are important but situated in the context of wider structures. Actors interact (struggle, have economic transactions, form alliances, exercise power, negotiate) within the constraints and opportunities of existing structures, while simultaneously acting upon and restructuring these systems. Another important point is that structures not only *constrain* but also *enable* action, that is, make rational action possible by providing coordination and stability.

To conceptualise coordination, it is common to use the term 'institutions'. But institutions are sometimes confused with (non-market) organisations. Hence, I prefer to use the general sociological concept of 'rules' instead. Although one can quarrel about terms and exact definitions, it is more important to look at the general phenomena they aim to describe, that is, coordination and structuration of activities. With regard to that aim, rules are similar to institutions.

Rules are not only formal rules, such as laws and regulations. Following Scott (1995) I distinguish three dimensions or 'pillars': regulative, normative and cognitive rules. The *regulative* dimension refers to explicit, formal rules, which constrain behaviour and regulate interactions, for example, government regulations, which structure the economic process. It is about rewards and punishments backed up with sanctions (for example, by the police or the courts). Institutional economists tend to highlight these formal and regulative rules (for example, Hodgson, 1998). North (1990), for instance, highlights rules which structure economic processes at the national level (for example, property rights, contracts, patent laws, tax structures, trade laws, legal systems). *Normative* rules are often highlighted by traditional sociologists (for example, Durkheim, 1949; Parsons, 1937). These rules confer values, norms, role expectations, duties, rights and responsibilities. Sociologists argue that such rules are internalised through socialisation processes. *Cognitive* rules constitute the nature of reality and the frames through which meaning or sense is made. Symbols (words, concepts, myths, signs, gestures) have their effect by shaping the meanings we attribute to objects and activities. Social and cognitive psychologists have focused on the limited cognitive capacities of human beings and how individuals use schemas, frames, cognitive frameworks or belief systems to select and process information (for example, Simon, 1957). Evolutionary economists and sociologists of technology have highlighted cognitive routines, search heuristics, exemplars, technological paradigms and technological frames of engineers in firms and technical communities (for example, Dosi, 1982; Nelson and Winter, 1982; Bijker, 1995). Table 1.1 briefly indicates the differences between these types of rules.

Rules do not exist as single autonomous entities. Instead, they are linked together and organised into rule systems. Such systems may be purely private rule or 'personality systems' or they may be collectively shared systems. The latter case refers to social rule systems. These systems, which structure and regulate social transactions and which are backed by social sanctions and networks of control, are referred to as 'rule regimes' (Burns and Flam, 1987: 13). I understand regimes as semi-coherent sets of rules, which are linked together. It is difficult to change one rule, without altering others. The alignment between rules gives a regime stability.

Social Groups Related to Socio-technical Systems

The elements of socio-technical systems do not function on their own, but are actively created and maintained by human actors, embedded in social groups. In modern societies many specialised social groups are related to elements of socio-technical systems. Figure 1.4 gives a schematic representation.[2]

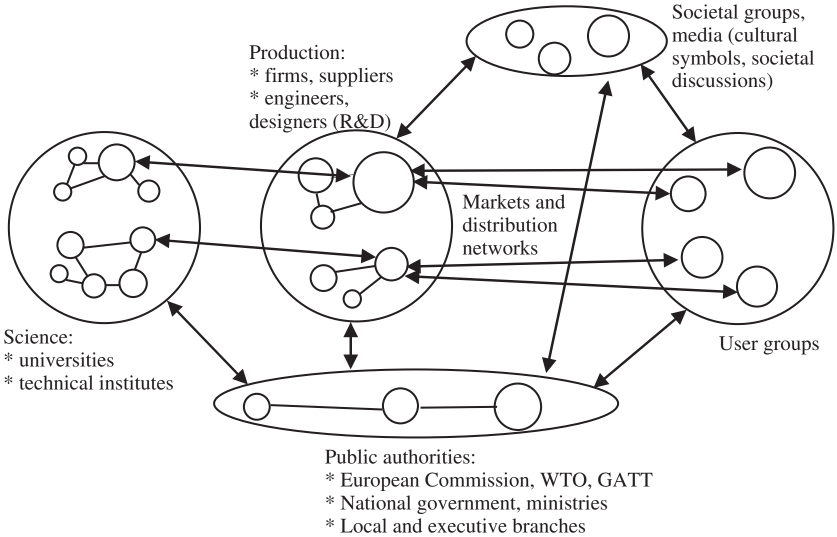
The configuration of social groups is the outcome of historical differentiation processes. Over time, social groups have specialised and differentiated, leading to more fine-grained social networks. The chains of social

Table 1.1 Varying emphasis: three kinds of rules/institutions

	Regulative	Normative	Cognitive
Examples	Formal rules, laws, sanctions, incentive structures, reward and cost structures, governance systems, power systems, protocols, standards, procedures	Values, norms, role expectations, authority systems, duty, codes of conduct	Priorities, problem agendas, beliefs, bodies of knowledge (paradigms), models of reality, categories, classifications, jargon/language, search heuristics
Basis of compliance	Expedience	Social obligation	Taken for granted
Mechanisms	Coercive (force, punishments)	Normative pressure (social sanctions such as 'shaming')	Mimetic, learning, imitation
Logic	Instrumentality (creating stability, 'rules of the game')	Appropriateness, becoming part of the group ('how we do things')	Orthodoxy (shared ideas, concepts)
Basis of legitimacy	Legally sanctioned	Morally governed	Culturally supported, conceptually correct

Source: Scott (1995: 35, 52).

groups have lengthened over time (Elias, 1982). In the Middle Ages, production and consumption were situated closely together. Knowledge, capital and labour were often located in the same producer (for example, a blacksmith). In the last two centuries production and consumption have increasingly grown apart, because of efficient, low-cost transportation systems and because of mass-production methods (Beniger, 1986). The lengthening of networks led to an increase in social groups: distribution involved an increasing number of social groups (for example, merchants, wholesalers, retailers, chain stores); techno-scientific knowledge has become more distributed over a widening range of actors (universities, laboratories, consultancies, R&D units in firms). The production of cultural and symbolic meanings involves an increasing range of mass media (newspapers, magazines, radio, TV, Internet), especially in the twentieth century. This dynamic of specialisation and differentiation means that it is not possible to define boundaries of social



Source: Geels (2002: 1260).

Figure 1.4 Social groups which (re)produce socio-technical systems

networks once and for all. Relationships between social groups shift over time and new groups emerge. In the electricity sector, for instance, liberalisation gave rise to electricity traders at spot markets as an entirely new group. This example also points to the fact that the precise configuration of social groups differs between sectors. The social network in transport systems looks and functions differently from that in electricity systems. This means that boundary definition is more an empirical issue than a theoretical one.

These social groups have relative autonomy and each social group has its distinctive features. Members share particular perceptions, problem agendas, norms, preferences and so on. They share a particular language ('jargon'), tell similar stories of their past and future, meet each other at particular fora, read the same journals and have professional associations and lobby clubs. In short, there is coordination *within* groups because members share cognitive, formal and normative rules. This means that different regimes co-exist, for example, technological regimes (design, production), policy regimes, science regimes, socio-cultural regimes and user/market regimes. Some examples of rules in technological regimes are search heuristics, guiding principles, promises and expectations, design criteria, functional product requirements, missions formulated by marketing departments, problem-solving strategies, technical models and tacit knowledge. Examples of rules in science regimes include Kuhnian paradigms, publication patterns, peer review procedures, criteria and methods of knowledge production, and exemplars. Examples of rules in policy regimes are policy goals, regulations and laws (for example, emission norms), R&D subsidy programmes, general guiding principles (for example, liberalisation, decentralisation). Examples with regard to users and markets are user practices and preferences, selection criteria, user competencies, procedures and practices for user feedback, and market institutions. Examples of rules in socio-cultural regimes are cultural values and beliefs (technological progress, enlightenment, freedom), myths about technological change (for example, logical progress), cultural symbols and fears (for example, Frankenstein, modernity).

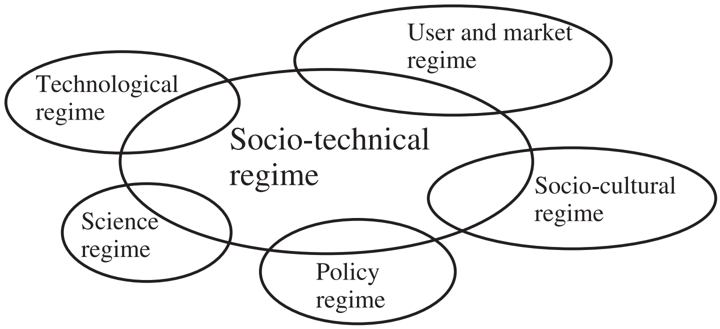
But different groups also interact with one another, and form networks with mutual dependencies. Although groups have their own characteristics, they are also interdependent and interpenetrative (Stankiewicz, 1992). Because of the interdependence, activities of social groups are aligned to one another. This means that there is also inter-group coordination, represented by linkages between rules of different regimes. The search heuristics of engineers are usually linked to user representations formulated by marketing departments. In stable markets, these user representations are linked with user preferences. Search heuristics are also linked to product specifications, which can be linked to formal regulations (for example, emission standards). This means that there are linkages between regimes, and this helps to explain the alignment of activities between different groups. To understand this meta-coordination I propose the concept of 'socio-technical regimes'. Such regimes do not encompass the entirety of other regimes, but only refer to those rules that are aligned to one another (see Figure 1.5). It indicates that different regimes have relative autonomy on the one hand, but are interdependent on the other.

Interrelatedness between Three Analytic Dimensions

The arguments above lead to an analytic framework of three interrelated dimensions. Socio-technical systems are maintained and reproduced by human actors and social groups. Human action is coordinated by rules and institutions. This means that socio-technical systems, actors and regime rules and institutions can be analytically distinguished (see Figure 1.6). However, they are always interrelated in practice.

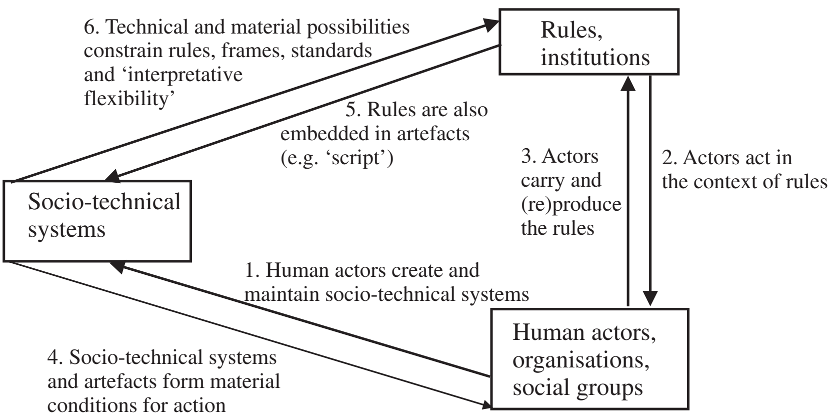
Between the three dimensions, there are six kinds of interaction:

1. Actors reproduce the elements and linkages in socio-technical systems in their activities. This point has been made and empirically illustrated in approaches in sociology of technology.



Source: Geels (2004a).

Figure 1.5 Meta-coordination through socio-technical regimes



Source: Geels, (2004a).

Figure 1.6 Three interrelated analytic dimensions

2. Actors, their perceptions and activities are embedded in regimes, rules and institutions.
3. But actors are not determined by rules and regimes. They implement and (re)produce the rules in their activities.
4. While this 'duality of structure' has been well conceptualised in sociology, this discipline almost entirely neglects the material nature of modern societies. Human beings in modern societies do not live in a biotope, but in a 'technotope'. Actor–network theorists emphasise that

we are surrounded by technologies and material contexts, ranging from buildings, roads, elevators, appliances and so on. These technologies are not only neutral instruments, but also shape our perceptions, behavioural patterns and activities. Socio-technical systems thus form a structuring context for human action. The difference between baboons and human beings is not just that the latter have more rules which structure social interactions, but also that they interact in a technical context (Strum and Latour, 1999).

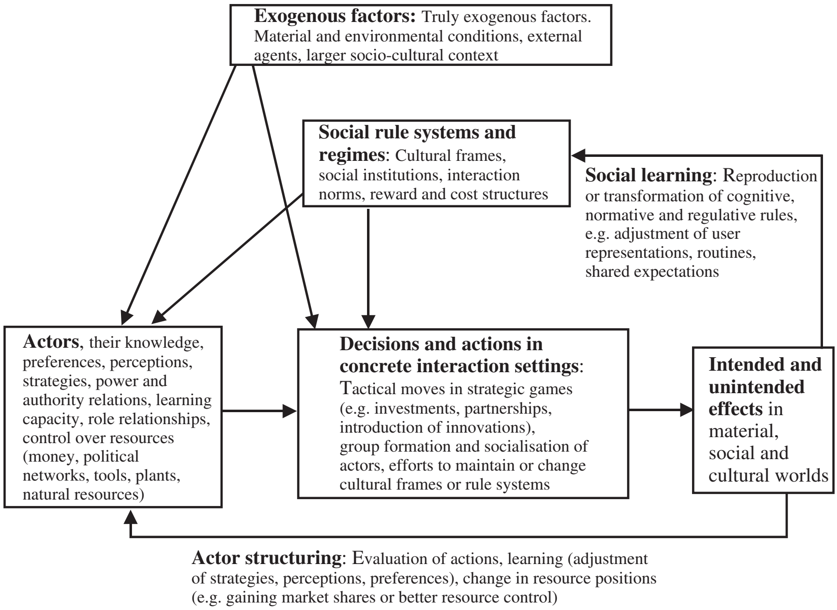
5. Rules are not just carried inside an actor's head, but can also be embedded in artefacts and practices. Notions of how rules are embedded in artefacts can be found in the philosophy of technology. For instance, Langdon Winner (1980) advanced the notion that technologies could have political effects built into them. Winner described the example of Moses's bridges on Long Island, New York, which were built very low, so that only automobiles could pass under them, not buses: 'Poor people and blacks, who normally used public transit, were kept off the roads because the twelve-foot buses could not get through the overpasses. One consequence was to limit access to Jones Beach, Moses's widely acclaimed public park' (ibid.: 128). Actor–network theorists such as Akrich and Latour (Akrich, 1992; Akrich and Latour 1992; Latour 1992) introduced the notion of a 'script' of an artefact to capture how technological objects enable or constrain human relations as well as relationships between people and things: 'Like a film script, technical objects define a framework of action together with the actors and the space in which they are supposed to act' (Akrich 1992: 208).
6. Technologies have a certain 'hardness' or obduracy, which has to do with their material nature, but also with economic aspects (for example, sunk costs). Because of this hardness, technologies and material arrangements may be harder to change than rules or laws. They may even give social relationships more durability (Latour, 1991). This hardness also implies that artefacts cannot entirely be shaped at will. There are limits to the interpretative flexibility of artefacts. Technical possibilities and scientific laws constrain the degree to which interpretations can be made. Next to social shaping, there is also technical shaping (Vincenti, 1995; Molina, 1999).

These six interactions make up the basis of a socio-technical analysis, which takes into account on the one hand that social actions influence technological change, but on the other that technology influences society. Social systems thus adapt themselves to changing technical systems, as well as the converse. Social shaping of technology is accompanied by technical shaping of society.

To make the analytical framework more dynamic, the interactions between the three dimensions can be further conceptualised: (a) dynamic interactions between rules and actors, (b) dynamic interactions between actors and systems, and (c) dynamic interactions between rules and systems.

Dynamic interaction between rule regimes and actors

As members of social groups, actors share a set of rules or a regime, which guide their actions. These rules are the outcome of earlier (inter)actions. Social actors knowledgeably and actively use, interpret and implement rule systems. They also creatively reform and transform them. Rules are implemented and (re)produced in social activities which take place in concrete interaction settings (local practices). Through implementing the shared rule systems, the members of communities generate patterns of activity, which are similar across different local practices. While there is similarity to some degree, there is also variety between group members. Members also have private rule systems, somewhat different strategies, different resources and so on. As a result, there may be variation between local practices within a shared social rule system. The strategies, interests, preferences and so on are not fixed, but change over time as a result of social action. Actors act and interact with each other in concrete settings or local practices. For instance, firms make strategic investment decisions, public authorities make new policy plans and regulations, and so on. The aim of these actions is usually to improve their situation and control of resources (for example, earning money, improve market position), that is, it is motivated by self-interest. Enactment of social rules in (inter)action usually has effects on the physical, institutional and cultural conditions of action, some of which will be unintended. Some effects will directly influence actors, for example, their resources, market shares and money. These direct effects are called 'actor structuring'. This may involve individual learning when specific actors (for example, firms) evaluate their actions, learn, and adjust their strategies, aims, preferences and so on. Other effects influence the shared rule system (for example, perceptions of who the users are, what they want, which technical recipes work best) and are called 'social learning', because they take place at the level of the entire group. This takes place through imitation[3] (firms imitate routines from successful firms) or through the exchange of experiences, for example, articulation of problem agendas, best practices at conferences, through specialised journal or professional societies and branch organisations. Through the effects of social interaction, social rule systems as well as social agents are maintained and changed. Figure 1.7 gives an impression of these basic dynamics. The figure also includes exogenous factors which conditionally structure actors, social action and system development, but which are not influenced by them



Source: Adapted from Burns and Flam (1987: 4); courtesy of Sage Publications.

Figure 1.7 Actor–rule system dynamics

(Burns and Flam, 1987: 3). These exogenous factors may change over time and impact on social rule systems, causing internal restructuring.

Figure 1.7 includes two feedback loops, an upper one (social learning) and a bottom one (actor structuring). The upper loop represents sociological and institutional dynamics, and can best be applied on longer time-scales (years, decades). For example, it is often years before government policies have any substantial effects at the level of systems. Likewise the articulation of new user preferences or new technical search heuristics may take years, because it occurs in small incremental steps, and often involves experiments and setbacks. Examples are the public acceptance of Walkmans (Du Gay et al., 1997) or the development of wind turbines (Garud and Karnøe, 2003). The bottom loop represents direct interactions between actors, which affect their resource positions (for example, money, market share) and network relationships. This includes dynamics which are emphasised in business studies and industrial economics, for example, strategic games in markets, power struggles, strategic coalitions and the innovation race. The time-scale of this bottom loop is usually shorter (for example, months, years).

Figure 1.7 thus combines and positions sociological and economic analyses. It shows that economic actions and transactions are embedded in sociological frames (see also Granovetter, 1985). Economic actions and transactions are also embedded in socio-technical networks, as argued by the earlier conceptualisation of technology. Markets for economic transaction are not simply there, but have to be created, especially in the case of radically new technologies. Market creation in such circumstances involves innovating firms in novel 'constituency building', as much as in novel technical invention (Green, 1992). Such constituencies involve groups within firms, but also external groups, for example, regulatory agencies, standard-setting bodies and external sources of finance. New technologies, users and markets are the outcome of a process of linking elements together. As the linkages between heterogeneous elements become more stable, markets emerge for particular technologies. Economic calculation is possible because socio-technical networks and meanings have stabilised:

> If agents can calculate their decisions, it is because they are entangled in a web of relations and connections; they do not have to open up the world because they *contain* their world. Agents are actor-worlds. . . . The agents, their dimensions, and what they are and do, all depend on the morphology of the relations in which they are involved. (Callon, 1999: 185)

Markets are made up and constitued by rules. Firms and customers need rules to enforce reliability and create stable playing fields (rules of the game):

> In order for commerce to grow in any uncharted territory there need to be rules. Not regulation necessarily, or even governments, just rules. There need to be property rights, for example, and some sense of contracts. In higher technology areas there need to be rules for intellectual property (who owns the operating system? Under what terms?) and provisions for standardisation (how do different products work together? Which technical platform becomes the norm?). Without these rules, commerce may still emerge, but it will not flourish. (Spar, 2001: xviii)

The aim of this conceptualisation is not to argue for the ultimate primacy of sociology, but to develop a dynamic framework, where economic activities and processes are on the one hand structured, but on the other hand influence and transform the sociological structures in which they are embedded. This basic conceptualisation does not mean that economic processes are not important. System innovations cannot be understood by only looking at rules, perceptions and linkages. Economic aims, strategies and investments are obviously important. But they are embedded in sociological rule systems and socio-technical networks. For short-term analyses the sociological structures may be assumed to be relatively constant, providing a frame for R&D strategies, strategic games and so on. For longer-term analyses (for example, changes from one socio-technical system to another) the sociological loop also needs to be included, and attention should be paid to social learning and institutional change.

Dynamic interactions between actors and systems: making moves in games

On the one hand socio-technical systems are maintained and changed by activities of actors; on the other they form a context for actions. We can understand these actions as moves in a game, of which the rules alter somewhat while the game is being played. Economic processes are embedded in sociological processes, but are not entirely determined by them. Within rules and regimes there is plenty of room for intelligent interpretation, strategic manoeuvring and so on. Institutional economists coined the notion of 'rules of the game'. This notion can be widened to mean that rules and regimes constitute a game, which is played out by actors, firms, public authorities, users, scientists, suppliers and so on. The different social groups each have their own perceptions, preferences, aims, strategies, resources and so on. Actors within these groups act to achieve their aims, increase their resource positions and so on. Their actions and interactions can be seen as an ongoing game in which they react to each other. The feedback loops in Figure 1.7 indicate that there can be multiple development rounds. In each round actors make 'moves', that is, they do something, such as make investment decisions about R&D directions, introduce new technologies in the market, develop new regulations and propose new scientific hypotheses. These actions maintain or change aspects of socio-technical systems. The dynamic is game-like because actors react to each other's moves. These games may be *within* groups, for example, firms who play strategic games between each other to gain competitive advantage. There may also be games *between* groups, for example, between an industry and public authorities. For instance, public authorities want to stimulate the environmental performance of cars, but they do not know exactly which regulations and emission standards are feasible. The car industry wants to prevent very strict regulations, but it also wants to demonstrate its good will to the public authorities ('with this new clean car, we are doing the best we can'). If one company opts for a strategy to introduce an even cleaner car, this changes the game, because it allows the authorities to introduce stricter rules to force other companies to do the same. With the stricter emission rules, the game has changed (somewhat). The added value of this conceptualisation (compared to institutional economists) is that the 'rules of the game' are not fixed, but may change during the game, over successive development rounds. It also shows how socio-technical systems change because of activities and (strategic) games between actors. The notion of 'playing games' also highlights that social (inter)action in the context of regimes is not necessarily harmonious. Different actors do not have equal power or strength. They have unequal resources (for example, money, knowledge, tools) and opportunities to realise their purposes and interest, and influence social rules. The rule-system framework leaves room for conflict and power struggles. After all, there is something at stake in the games that are being played.

Dynamic interactions between rules and systems

The different social groups are partly autonomous, operating in their own regime. The cognitive, formal and normative rules extend across the playing fields in each regime. Actors in each regime make moves in games as it were. They bring an improved product to the market, develop a new law or regulation, make changes to the infrastructure and write scientific articles with new knowledge claims. The moves thus have some tangible effects on elements of socio-technical systems. Moves react and build upon each other. Effects of moves are evaluated, and may lead to changes in rules, perceptions and strategies. The accumulation of moves leads to ongoing processes or trajectories in regimes. In stable regimes these trajectories tend to go in particular directions (see Figure 1.8). Each small arrow in Figure 1.8 represents a move with some tangible effect on elements of socio-technical systems. Each of these moves can be analysed using Burn and Flam's actor–rule interaction in Figure 1.7 (moves occur in the context of particular rules; the effects can lead to changes or reproduction of rules). Because the regimes interpenetrate, moves in one regime may influence moves in another. For example, changes in user preferences and markets may lead to changes in technology, which, in turn, lead to changes in

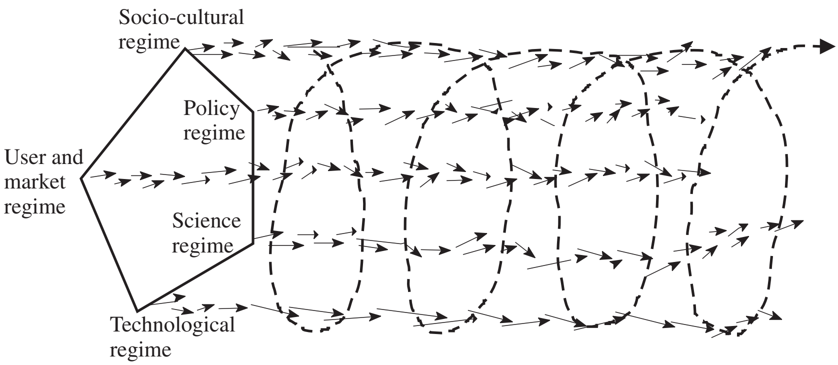


Figure 1.8 Alignment of ongoing processes in a socio-technical regime

regulation. As the technology diffuses it may trigger societal debates about its possible effects, which triggers more regulation. This regulation may change the direction of technological change and so on. Such chains of moves lead to a kind of socio-technical leapfrog dynamic, in which the different regimes co-evolve.

1.4 RESEARCH STRATEGY AND CASE SELECTION

Research Strategy

The conceptualisations above do not yet answer how system innovations come about. But they do provide an understanding of the structure and functioning of socio-technical systems. They also show that dynamics in such socio-technical systems are complex and multi-causal. These considerations should guide the choice of a research strategy.

This book uses case studies as a research strategy. The phenomenon of system innovation is a process that unfolds over time, involving major changes in socio-technical systems. As such, system innovations are a complex phenomenon, consisting of many interacting elements and linkages. Investigations of this kind of phenomenon require a research strategy that is both rich in context and can track complex developments over time. Yin (1994) argues that case studies are the only methodology to study a phenomenon in relation to its real-life context and to understand causal links in complex situations where many variables interact. Case studies are well suited to tell rich stories in terms of dynamics and interacting processes. The focus is on the interlocking of multiple processes and activities, not on linear cause-and-effect relationships.

The main research question is: how do technological transitions come about? Case studies have a distinct advantage over other research strategies when a 'how' question is being asked: 'This is because such questions deal with operational links needing to be traced over time, rather than mere frequencies or incidence' (ibid.: 6).

There are different case-study designs, for example, single or multiple cases, holistic or embedded. This book uses a multiple case-study research design, which is better than using just one case study. If the conceptual perspective holds in multiple cases, it becomes more robust (empirical replication). Multiple case studies also have the advantage that contrasts and comparisons can be made, which may lead to refinement of insights.

The aim of the case studies is to test the conceptual perspective. This can be done through pattern matching (ibid.: 106), which compares an empirically based pattern with a predicted one. If the patterns coincide, the results can help to strengthen the internal validity of the propositions. This means that the question of whether the empirical case studies have a good 'fit' with the hypotheses has to be assessed in the conclusions.

Case Selection

I have chosen the following three case studies, focusing on the leading countries:

1. The transition from sailing ships to steamships in British oceanic transport (1780–1900).
2. The transition from horse-drawn carriages to automobiles in American urban passenger transport (1860–1930).
3. The transition in aviation from piston engine aircraft to jetliners in American aviation (1926–1975).

I have chosen these case studies because they are critical or revelatory (ibid.: 40). They have been well studied by authors from other disciplines, often following a simple technical substitution storyline. If I can show that their storyline is insufficient, this shows the added value of my perspective.

Another reason for this selection is that they are historical cases where the transition is complete, and on which historical descriptions have stabilised. Also, sufficient secondary literature is available to make reconstructions manageable.

All three case studies come from the domain of transportation (water, land, air). I have chosen this domain, because all elements from the socio-technical system play a prominent role, including infrastructure and culture, which may be less important for other domains (for example, materials supply). Furthermore, technology and society have strong interactions in this domain. Wars had a great influence on transport developments, because ships, cars and planes acted as defence and attack weapons, to transport troops and for logistics. Transport was also influenced by macro-societal developments such as liberalisation of trade, industrialisation and emigration, and by cultural values such as speed and freedom. But transport innovations also impacted on society. Urban life was greatly transformed by the automobile, leading to an expansion of roads, segregation of public space and suburbanisation. In rural areas, the car also eroded institutions such as the school, church, shops and medical facilities. The transition to jetliners was accompanied by a decrease in travel fares, which strongly stimulated worldwide travel and tourism, making the world a smaller place. The transition to steamships also led (eventually) to lower fares, stimulating emigration and worldwide trade. Steamships made possible the large-scale import of grain in Europe, stimulating feeding patterns but threatening local agriculture. In sum, the transport domain is a promising area to explore the dynamics of system innovations.

Data Collection

Case studies are a laborious method, because it is time-consuming to collect empirical materials and arrange them in plausible storylines. Because I chose multiple case studies over long time periods it is not possible to study each case in detail. The studies are not intended primarily to unveil new historical facts, but to illustrate the conceptual perspective. Hence I shall use secondary, published material to describe the case studies.

As an additional source of information, I consulted sectoral experts – aviation: Dr Marc Dierikx, *Instituut voor Nederlandse Geschiedenis* (Institute for Dutch History), The Hague; horse transportation and automobiles: Dr Gijs Mom, *Stichting Historie der Techniek* (Foundation for the History of Technology), University of Technology, Eindhoven; and shipping: Dr Hugo van Driel, Department of History and Art Sciences, Centre for Business History, Erasmus University, Rotterdam, and Professor Frank Broeze (e-mail contact), Department of History, University of Western Australia, Crawley. First, the experts generated a list of relevant references and, using these as a starting point led to the discovery of many more. Second, I drafted an empirical story which was sent to the experts, who were then interviewed about possible gaps, factual errors or wrong interpretations. On the basis of their comments the early draft was rewritten.

1.5 OUTLINE OF THE BOOK

The first part of the research is explorative and conceptual. Chapter 2 offers a wide-ranging, multidisciplinary critical literature review. It generates relevant insights with regard to aspects of system innovations, 'bits and pieces', which can be used as building blocks for a coherent perspective. Chapter 3 takes an existing framework, the multi-level perspective on regime shifts, as starting point. This perspective has some weaknesses for understanding system innovations. Building blocks from Chapter 2 will be used to address these weaknesses and suggest an improved multi-level perspective to understand system innovations and technological transitions. Conceptual propositions are advanced as answers to the research questions.

The second part of the research tests these conceptual propositions, with the help of three case studies: Chapter 4 describes and analyses the transition from sailing ships to steamships; Chapter 5 describes and analyses the transition from horse-drawn carriages to automobiles; and Chapter 6 describes and analyses the transition from piston engine aircraft to jetliners. These case studies address the entire process of system innovation in its complexity. Chapter 7 draws conclusions about the conceptual propositions by matching them with the case studies. A further reflection on the case studies and the improved multi-level perspective leads to further refinements, in the form of a proposal of different transition routes. The research question about mechanisms is addressed in a separate section. The book ends with some critical reflections on the type of knowledge claims and generalisability of the findings.

NOTES

1. The large technical systems theory (for example, Hughes, 1987) is somewhat of an exception. In his analysis of the development of LTS, Hughes distinguished different phases, and identified mechanisms relevant for these phases (for example, load factor, diversification, momentum).
2. Figure 1.4 can be made more complex by zooming in on actors *within* groups and linkages *between* groups. Thus we can also find professional societies, trade associations, distributors, industry consortia, consulting companies, semi-public government agencies, private research institutes and standard-setting bodies.
3. See also Nelson and Winter (1982: 135) according to whom 'imitation is an important mechanism by which routines come to organise a larger fraction of the total activity of the system', thus playing a role in the emergence of technological regimes.

2. Building blocks from the literature

2.1 INTRODUCTION

There is no specific literature about system innovation and technological transition, as delineated in Chapter 1. But there is a wide range of literature on particular aspects of system innovations. To structure the literature review, the focus in subsequent sections is on three aspects: (a) system approaches, (b) technological change and (c) co-evolution. System approaches are dealt with, because the topic of the book is system innovation (Section 2.2). Literature on technological change is taken into account, because technology is taken as an entry point (Section 2.3). This will include sociological, socio-technical and economic literature. Co-evolution is dealt with because it is a crucial process in system innovation (Section 2.4).

The aim of this chapter is to generate insights and concepts that can be used as building blocks for the construction of a conceptual perspective in Chapter 3. This means that the discussion of the literature will be somewhat eclectic.

Although this book aims to develop a co-evolutionary and socio-technical perspective on system innovations, the literature review is relatively wide, but not exhaustive. In Chapter 1 (Section 1.3) it was argued that sociological frames and socio-technical systems provide a structuring context for economic (trans)actions. But although economic processes are embedded in sociological and socio-technical contexts, they are important to understand system innovations. Hence, building blocks can be found in a wide range of literature.

2.2 SYSTEM APPROACHES

There are several approaches that have technical systems as their topic of analysis, for example, large technical systems theory, and complex products and systems. There is also literature that takes systems as an analytical approach, for example, sectoral systems of innovation, and the technological systems approach. The systems approach means that linkages between elements are analysed, but the subject of the analysis is social networks involved in innovation, rather than technical systems.

Large Technical Systems Theory (LTS)

Large technical systems theory is a conceptual approach within science and technology studies (Hughes, 1983, 1986, 1987; Mayntz and Hughes, 1988; La Porte, 1991; Coutard, 1999). LTS research focuses on a particular kind of technology, consisting of interrelated technologies, often with infrastructural networks, which stretch geographical areas (for example, electricity systems, telephone networks, railroad systems and the Internet). The LTS approach also developed a particular socio-technical mode of analysis, looking at seamless webs. The term 'seamless web' highlights the heterogeneous character of LTS. LTS consist not only of technologies, but also of people, firms, texts, contracts and legislation. All these aspects need to be moulded and aligned if the socio-technical configuration as a whole is to work. System builders (like Thomas Edison in the case of electricity networks) travel between domains such as economics, politics, technology, applied scientific research and aspects of social change, weaving an emergent web, which links the bits and pieces together. The emphasis is thus on heterogeneous linkages and the work of system builders.

LTS research has focused mainly on the emergence and development of large technical systems, which is loosely described as a life cycle. Hughes (1987) distinguished five phases in the development of LTS: (a) invention, (b) development, (c) innovation, (d) growth, competition and consolidation, and (e) momentum. The different phases overlap and backtrack. After invention, development and innovation, there is more invention.

Inventions occur during all phases. Inventors work in particular communities, publishing their patented inventions in technical journals: 'These publications inform the inventive community about the location of inventive activity. This alerted the inventive community about problems that needed attention, for rarely was a patent or invention the ultimate solution to a problem' (ibid.: 60). Hughes thus argues for continuity of invention, in terms of gradual developments in problem agendas, search directions and eventual solutions.

A radical invention needs to be further developed. Inventor–entrepreneurs need to attract sponsors. The task is to create a 'configuration that works', not only in the technical sense, but also economically, socially and politically. The design stage is characterised by flexibility of design, openness to social, cultural and political influences, competition between different options, turbulence and uncertainty.

The phase of innovation is reached when the physical components are made to work in a commercial configuration consisting of manufacturing, sales and service facilities. As the organisational complexity increases, manager–entrepreneurs become more important.

In the growth phase, economies of scale are important, but there are also three additional concepts: load factor, economic mix and reverse salient. The 'load factor' is the ratio of average output to the maximum output during a specified period. The task of managers is to acquire high load factors, thus best utilising the system. They strive for diversified demands, so that the peak of one kind of demand occurs when another demand is low. Managers also seek a good 'economic mix', that is, a variety of input sources for the system, which they can optimise according to cost and availability. 'Reverse salients' are internal problems which occur as a system expands. They are components in the system that have fallen behind or are out of phase with the others. Reverse salients guide technological development in systems, when they are defined as a set of critical problems.

After prolonged growth, systems acquire momentum. This stage is characterised by the wide embedding of an LTS in society. Technical systems have a mass of technical and organisational components; they possess direction or goals; and they display a rate of growth suggesting velocity. The early Hughes (1987) mainly emphasised social and organisational elements to understand momentum. All kinds of organisations have vested interests in the continuation and further development of the system, actively working and lobbying for this, and thus creating a momentum. But Hughes (1994) more explicitly emphasises the importance of technical elements for momentum. People adapt their lifestyles to artefacts, new infrastructures are created and industrial supply chains emerge, making part of the economic system dependent on the artefact. Because of all these linkages, it becomes almost unthinkable for the technology to change in any substantial fashion. A *reversal* occurs as the technology shifts from flexibility to dynamic rigidity (Staudenmaier, 1989). The technology shifts roles from a possible social option, open to a host of different futures, to a culture-shaping and highly specified social force. It is not an autonomous force, however, because the momentum is actively created by the 'weight' of linkages.

Hughes says little about the decline of LTS. This suggests that systems persist for ever once they have reached the momentum stage. Staudenmaier (1989) suggested including a senility stage, which can be understood as a weakening of the linkages between technology and the wider context:

> As new political priorities, shifting demographics, changing tastes, ecological transformations, or competing actors come to the fore, the sweet fit between

> context and technology that characterized the momentum stage begins to unravel. (p. 155)

This idea that systems become unstable, when the linkages between the elements weaken, will be a useful building block in Chapter 3 to understand the change from one system to another. This kind of dynamic has not been studied much in the LTS literature. One exception is the edited book by Summerton (1994), which explores how momentum can be overcome and how systems change. In her introduction, Summerton argues that: 'Periods of stability in technical systems and networks are typically only provisional. Systems and networks are dynamic entities. They can seldom be black-boxed for good' (p. 5). Previously achieved closure is undone when the linkages between elements of the system begin to weaken. The weakening of linkages opens up the potential for transformation. Although Summerton says little about the dynamics of transformation, she notes several prevailing factors for the opening up of systems. One factor is the presence of underlying problems within the system ('reverse salient' in Hughes's terminology) and actors' view of these problems. A second factor relates to problems external to the system, so-called 'negative externalities', for example, environmental impacts, risks and concerns about safety. Public criticism of externalities is often expressed as consumer pressure to alleviate the problem. Consumer demands may lead to political or regulatory pressures to improve performance. A third factor is changing competitive conditions, which can trigger reconfiguration, for example, new innovations and markets. A fourth factor relates to political developments or contingencies in an even broader sense (for example, war or threat of war). Such factors can radically alter the shape and direction of LTS. A fifth factor relates to changes in cultural values and expectations, or broad political ideologies.

Some contributors in Summerton (1994) propose particular dynamics to understand changes in LTS. Von Meier (1994) sketches a possible future transformation of the electricity system, whereby new technologies are first introduced in the existing electricity system to deal with particular problems (for example, photovoltaic solar cells, wind turbines, solar thermal systems, gas turbines and fuel cells). In a later phase these technologies open up new possibilities, because more is learned about their technical and functional characteristics. When these technologies take on a dynamic of their own, they may lead to transformations on multiple dimensions, for example, changing technology base, grid operation and control, planning, legal dimensions and ownership. The dynamic has the characteristics of a Trojan horse, which is initially welcomed for particular purposes, but eventually has unforeseen transformational consequences. It also suggests that old and new technologies need not compete from the start. New technologies may be adopted in the existing system, gradually becoming emancipated and then taking on the existing technologies. These are relevant building blocks for Chapter 3.

Both the emergence of LTS and their transformation are complex processes, consisting of co-evolving elements and multiple social groups. Hence, it is difficult for anyone to oversee and control changes in LTS. For system innovations this means that they are usually not planned, but emerge when multiple dynamics link up:

> Retrospective studies of LTS show that they never develop according to the designs and projections of dominant actors: LTS evolve behind the backs of the system builders, as it were. It has been shown, too . . . that typically none of the agencies contained in LTS manage to form a somewhat complete picture of their workings. LTS seem to surpass the capacity for reflexive action of actors responsible for operating, regulating, managing and redesigning them in ways which, as social scientists, we understand poorly. (Joerges, 1988, p. 26)

Complex Products and Systems (CoPS)

The CoPS approach also has a particular unit of analysis, namely, complex products and systems. CoPS can be defined as high cost, technology-intensive, customised, capital goods, systems, networks, control units, software packages, constructs and services (Hobday et al., 2000). Some examples are flight simulation systems (Miller et al., 1995), air traffic control systems, aircraft engines, bridges, chemical plants, passenger aircraft, jet fighters, robotics equipment, supercomputers and submarines. CoPS tend to be made in small batches, sometimes in one-off projects. They are often made for particular purposes, described by demanding users. They tend to be customised, and the user is often closely involved in the innovation process. CoPS are defined as 'complex', because of the high number of specialised components involved, the breadth of knowledge and skills required and the degree of new knowledge involved in production. Because of the complexity and capital intensity, CoPS are usually not made by one firm, but by networks of firms. Much of the CoPS literature is about the management of innovation in CoPS. The typical organisational and management structure is the project (Hobday, 1998).

Technical configurations and CoPS can be understood as a *technical hierarchy*, consisting of materials, components, devices, subsystems and artefacts (Disco et al., 1992: 485). In CoPS, these hierarchies may be very elaborate. Figure 2.1 gives an example of a technical hierarchy for the automobile (even though automobiles are officially not considered as CoPS, because they are mass produced and standardised). The components and

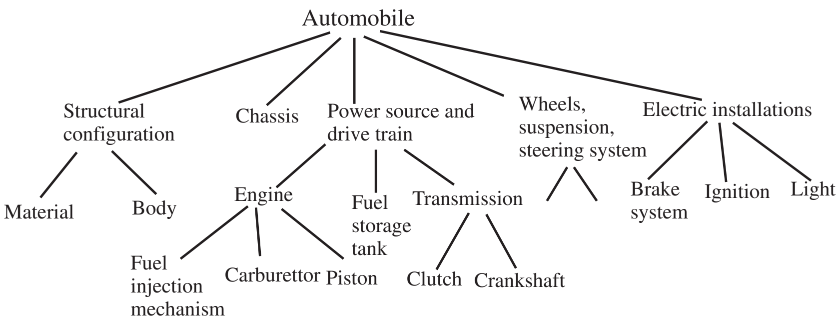


Figure 2.1 Technical hierarchy for the automobile

subsystems are interconnected in particular product architectures. The complex system is thus broken up into discrete pieces, which can communicate through standardised interfaces within a standardised architecture. This allows for the possibility of modular innovation (Sanchez and Mahoney, 1996; Langlois, 2002). As long as components and subsystems follow interface standards, data transmission protocols and component specifications, they can be modified and improved without affecting the working of other components and subsystems. This allows for distribution of labour. By concentrating on a single component or subsystem, each business unit or company can push deeper into its workings (Baldwin and Clark, 1997).

But modular innovation has its limits. Sometimes components are so drastically changed or improved that they require changes in other components or even change in the entire product architecture (Henderson and Clark, 1990). This is an interesting notion for my research questions. Substitution of a component may have wider *cascade dynamics* in the entire artefact. Changes may thus cascade from lower to higher levels in the technical hierarchy. Perhaps component changes can even cascade further upwards and influence the wider socio-technical system, thus contributing to system innovation. Within watches, for example, 'the emergence of quartz movement had cascading effects on all other watch subsystems as well as manufacturing processes' (Tushman and Murmann, 1998: 240).

A drawback for my research questions is that the CoPS approach has a strong focus on artefacts, and neglects its wider embeddedness. It has been argued that more attention should be paid to the interaction between CoPS and LTS (Geyer and Davies, 2000; Hobday et al., 2000). I would add that its interaction with socio-technical systems also deserves more attention.

Systems of Innovation and Technological (Innovation) Systems

The literature on systems of innovation and technological systems is different from LTS and CoPS approaches in the sense that it does not have systems as its topic of analysis. Instead, systems are used as a conceptual approach to study innovation. Linkages between elements are central in the systems of innovation approach. Because these elements are mainly social groups, the system approach is actually a social network approach. An important point, which it shares with the sociology of technology, is that innovation is not a single but a collective endeavour, emerging from the interaction of multiple actors. The scope of analysis is thus broadened from individual organisations (firms) to networks of organisations.

Systems of innovation can be defined on several levels. In the literature we find national systems of innovations (for example, Freeman, 1988; Lundvall, 1992; Nelson, 1993), regional systems of innovation (for example, Saxenian, 1994; Doloreux, 2002), sectoral innovations systems (for example, Breschi and Malerba, 1997; Malerba, 2002) and technological systems (Carlsson and Stankiewicz, 1991; Carlsson, 1997, 2002). Although these approaches have a different focus, they share an emphasis on linkages between elements, and they all see innovation as a co-evolutionary process.

For this book, sectoral innovations systems and technological systems have most to offer. The sectoral system of innovation approach (SSI) aims to provide a multidimensional, integrated and dynamic view of sectors. SSI can be defined as:

> a system (group) of firms active in developing and making a sector's products and in generating and utilizing a sector's technologies; such a system of firms is related in two different ways: through processes of interaction and cooperation in artefact-technology development and through processes of competition and selection in innovative and market activities. (Breschi and Malerba, 1997: 131)

This definition is about 'systems of firms', although non-firm organisations are mentioned in other definitions, for example, universities, financial institutions, central government and local authorities. This means that the focus is actually on social networks, for example, cooperation or knowledge flows. Another characteristic of the definition is that it does not look explicitly at the user side, although a selection environment is included. It is characteristic of most system of innovation approaches that they look mainly at the supply side of innovation. The user side is often taken for granted or narrowed down to a 'selection environment'.

An important insight is the emphasis on interlinkages between elements and the co-evolutionary process. For instance, Malerba (2002: 262) notes about sectoral systems: 'From the claim that the elements of a sectoral system are closely connected, it follows that their change over time results in a co-evolutionary process of its various elements. This process involves technology, demand, knowledge base, learning processes, firms, non-firm organisations and institutions'.

Institutions are recognised as important, although mainly in the form of formal institutions (standards, regulations, labour markets). Cognitive and normative institutions are lacking. This is a pity, because knowledge is seen as crucial for innovation and production. Learning processes receive much attention in SSI, not only the production of knowledge, but also its diffusion, for example, knowledge spillovers. That is where the relevance of social networks comes in. In well-functioning SSI, knowledge flows easily between the actors involved. Problems are well articulated and R&D activities targeted towards solving them. Different sectors can be characterised as to the main source of innovation, which can be R&D, specialised suppliers, demanding customers or scale-related problems in production (Pavitt, 1984).

The main focus in the SSI approach is on the *functioning* of systems. Sectors can be compared with regard to the source of innovation, network structures and degree of cooperation. Not much attention has been paid to the change from one system to another. In a recent discussion of sectoral systems of innovation, Malerba (2002: 259) noted that one of the key questions that need to be explored in-depth is: 'how do new sectoral systems emerge, and what is the link with the previous sectoral system?'. This question is addressed in this book.

A similar, but slightly different approach is that of technological (innovation) systems. A technological system is defined as:

> networks of agents interacting in a specific technology area under a particular institutional infrastructure to generate, diffuse and utilize technology. Technological systems are defined in terms of knowledge or competence flows rather than flows of ordinary goods and services. They consist of dynamic knowledge and competence networks. (Carlsson and Stankiewicz, 1991: 111)

It is clear from this definition that the focus is on social networks. The technological system approach differs from the systems of innovation approach in three ways. First, the former has a focus on a particular technology or set of technologies, while the latter looks at a sector, country or geographic region. Examples of technologies that have been studied are factory automation in Sweden (Carlsson and Jacobsson, 1993), renewable energy technologies (Jacobsson and Johnson, 2000) and biotechnology (Carlsson, 2002). These technologies are not systems, which makes the naming of the approach as technological systems a bit confusing. Second, the boundaries of technological systems do not necessarily coincide with national boundaries. Third, the technological systems approach places greater emphasis on microeconomic aspects, such as economic competence of firms and knowledge networks. Much attention has been paid to the (strategic) actions of individual firms. Firms differ in the information they have and how they use it; they differ in the resources they devote to advancing their knowledge base and their ability to learn from experience. The quality of organisations and managers matters. Firms differ in their ability to perceive opportunities and in their willingness to take risk (Carlsson and Stankiewicz, 1991: 100). These differences between firms lead to competition and strategic manoeuvring in particular technology areas.

The technological systems approach also pays attention to institutions, which can be 'hard' (legislation, capital market, educational system) or 'soft' (culture). Institutions greatly affect the specific path a technology takes. Jacobsson and Johnson (2000) distinguish three elements, which make up technological systems: actors and their competencies, networks, and institutions. Aspects of these three elements may hinder the emergence and diffusion of new technologies (Table 2.1). From these factors it is clear that there may be a large number of forces that block the formation of a new technological system. These forces may work independently but are likely to reinforce one another. This insight is an important building block for Chapter 3. However, I think that it is useful to make a closer distinction in Table 2.1 between factors related to the *new* technology and those relating to the *existing* system. Most factors in Table 2.1 relate to internal processes in the new technological system. But some indicate that the existing system may hinder or block the new one (established technology characterised by increasing returns, market control by incumbents). This distinction is also an important building block for Chapter 3.

The technology systems approach also identified crucial issues for a better understanding of the emergence of new systems. Jacobsson and Johnson identify three issues about which we need to know more. The first is the process of creating variety. Innovations in the existing system are likely to be incremental and cumulative, and more radical innovations may need to occur outside the existing system. But how does this occur? The second issue is the process of institutional change. Incumbent technological systems are often stabilised by existing institutions, for example, favourable legislation. The replacement of existing systems is thus likely to involve institutional change. But how does this occur? It may involve struggles in the political arena, and changes in policy instruments and regulations. The third issue is the emergence of prime actors or system builders. Existing systems are

Table 2.1 Examples of factors leading to a new technology being repelled

Actors and markets	Poorly articulated demand Established technology characterised by increasing returns Local search processes (cognitive routines which blind engineers to alternatives) Market control by incumbents
Networks	Poor connectivity Wrong guidance with respect to future markets
Institutions	Legislative failures Failures in the educational system Skewed capital market Underdeveloped organisational and political power of new entrants

Source: Jacobsson and Johnson (2000: 631).

also stabilised by incumbent actors and vested interests. To overcome this organisational inertia, promoters or product champions may be needed to raise awareness, undertake investments, provide legitimacy and diffuse the new technology. Such promoters may come from outside the system, when established actors have many vested interests. The insights regarding the three issues are important building blocks, but do not yet add up to an integrated understanding of how system changes come about. That is the challenge for Chapter 3. Some inspiration can be found in complex systems theory.

Complex Systems Theory, Hierarchy Theory and Evolution

Complexity theory or complex systems theory is an interdisciplinary research perspective, with intellectual roots in mathematics, linguistics, economics and biology. Complex systems theory is not (yet) so much a theory as a perspective for theorising and modelling system dynamics. At this point it is a heterogeneous ensemble of often exciting but not always polished mathematical observations (Morel and Ramanujam, 1999). Complexity researchers share a common fascination for the new insights into the dynamics of systems that are made possible by the advent of powerful computers. Interest in complex systems re-emerged in the 1990s with the advent of new mathematical computer modelling techniques, such as cellular automata, neural networks and genetic algorithms (Anderson, 1999). Although the bulk of the work has a mathematical focus, some conceptual work has also been done (for example, Potts, 2000;

Allen, 2001). That is the focus here. Empirical work has been lagging behind – there are few detailed, real-world case studies to back up complex systems theory.

An analysis of the importance of hierarchies has long been part of the study of complex systems (Simon, 1962; Pattee, 1973a). In application to the architecture of complex systems, 'hierarchy' simply means a set of Chinese boxes of a particular kind, usually consisting of a box enclosing a second box, which, in turn, encloses a third and so on. In hierarchic systems, the opening of any given box discloses not just one new box within, but a whole set of small boxes. The architecture thus has a tree-like structure. Simon (1973) distinguished four kinds of intertwined hierarchies (see Table 2.2). More could be added, for example, hierarchies in technology and knowledge. Complex systems should thus be seen as nested hierarchies. Complex systems at a certain level (for example, organisation, living organism) are thus built up from other complex systems (departments, tissues and organs) and so on. On the other hand, a system at a certain level is embedded in wider systems, which creates boundaries and constraints.

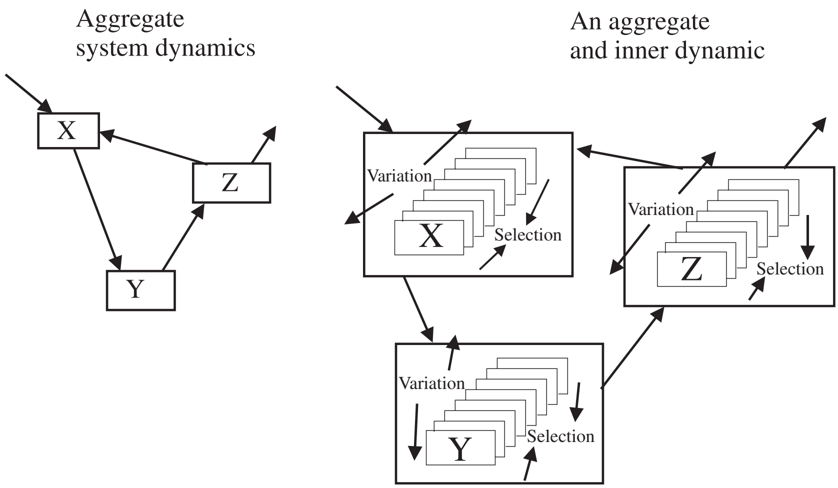
Table 2.2 Different nested hierarchies

Chemical/physical	Physiological	Biological	Organisational
Chemical substance	Living organism	Ecosystems	Human societies
Molecules	Tissues and organs	Species and populations	Groups of organisations (for example, sector, industry)
Atoms	Cells	Living organisms	Organisations
Nuclei and electrons	Macro-molecules	Genes and chromosomes	Departments in organisations
Elementary particles	Organic compounds	DNA	Individual human beings
	Chemical substances	Base pairs	Cognitive programmes in the central nervous system
		Chemical substances	Neural networks in the brain

Source: Simon (1973: 5).

Coupling and linkages are crucial in complex systems. There is horizontal coupling (between elements at the same level) and vertical coupling (between different levels). Linkages can vary on a continuum of strong and loose linkages. Loose horizontal coupling permits each element to operate dynamically independently of the others. Only the inputs and outputs it produces are relevant for the larger aspects of system behaviour. Loose coupling permits mutation (variation) and selection to go on in particular subsystems without requiring synchronous changes in all the other subsystems that make up the total system (Simon, 1973: 17). Loose coupling thus permits decomposition and modular innovation. Loose horizontal coupling basically means that the structure of the system is maintained, even while subsystems are improved. This is important to understand the *functioning* of systems, but does not say much about the *change* of systems. In the postscript of his book, Pattee (1973b: 145) identified the following unsolved problem: 'how do new levels of hierarchical organization arise? How do new structures appear without pre-existing prescriptions?'. He suspected that the answer had to be found in the interaction *between* different levels and that 'the theory of hierarchy must be closely related to the theory of evolution' (p. 150).

These ideas have been elaborated in the 1990s. Allen (2001) criticised system dynamics models, because they focus mainly on interaction between elements at the same level, represented by boxes and flows (written in mathematical equations). Such representations assume that the agents that are represented in such boxes have 'average' behaviour. But the crux of complex systems theory is that agents (elements or subsystems which make up systems) are characterised by microdiversity in behaviour. Individual agents are at different locations and have different experiences. Instead of average behaviour, differences between individual agents and local contexts lead to variation in behaviour. This difference between system dynamics and complex systems is represented in Figure 2.2. While the agents create variation, selection occurs at the level of the system: 'The essential process that therefore underlies complexity is the creative dialogue between the underlying biological and cultural micro-diversity that is exploring behaviour space, and the selective response of the rest of the system' (ibid.: 153). This means that the lower level generates variety, while the higher level exerts selective pressure. The consequence is that multiple variation and selection process occur in a nested hierarchy, an idea already advanced by Campbell (1974). Mokyr (2000: 62) applied this idea to technological change, arguing for multiple selection processes occurring simultaneously: (a) the market selects products produced by firms; (b) firms select different design options and cognitive routines; (c) firms and the routines they carry are selected through their success in the market; and (d) engineers and designs make selective choices from the wider scientific knowledge base.



Source: Allen (2001: 154); courtesy of World Scientific Publishing.

Figure 2.2 Mechanical system dynamics versus 'inner' level of micro diversity

The ideas about hierarchy and variation selection can be used to understand major changes in complex systems. Holling (2001: 390) argued that each level in a system is allowed to operate at its own pace, protected from above by slower, larger levels, but invigorated from below by faster, smaller cycles of innovation. In stable systems, the processes are robust and interlinked. Variations from below are dampened, leading to stability, but stable systems may be disturbed by outside events or processes at higher levels. An ecosystem, for example, may be disturbed by wind, fire, disease or insects. When such events disturb the system, this creates a window of opportunity for the breakthrough of innovations from the lower level. Such a breakthrough may then cascade and trigger further changes. Major system changes are thus understood as the result of linkages between processes at different levels. This forms a crucial building block for Chapter 3. A major challenge will be to make this abstract change model operational for socio-technical systems and use it for empirical case studies.

2.3 TECHNOLOGICAL CHANGE

Because I have taken technology as an entry point to study system innovations, this section takes stock of insights in different disciplines about the process of technological change. It is common to distinguish four kinds of processes or phases in technological change: (a) the emergence of new technologies (invention), (b) their linking to markets (innovation), (c) diffusion of new technologies and (d) replacement of old by new technology. A great deal of literature has dealt with these processes and phases. This section discusses three different approaches: sociological, socio-technical and economic. The literature in each approach highlights different aspects, starts with different assumptions, has different levels of analysis, or focuses on different phases. Hence, they all offer different insights and building blocks.

Sociological Approaches

In sociological literature, the focus is on activities, perceptions and cognitions of actors, which change over time as actors learn and interact. Technologies are seen as heterogeneous configurations, which need to be linked together. Co-evolution is thus obvious. The focus is often on the emergence of new technologies (invention), and their linking to markets (innovation). Less is said about diffusion of innovations and replacement of old by new technologies. This means the fluid phase takes centre stage, in which stabilisation occurs. New innovations stabilise as shared meanings and interpretations emerge. Some approaches argue that the 'new grows out of the old', and that new technologies are seen first with concepts related to the old technology. Only gradually do new technologies grow into their own, with specific design features and functionalities.

Social construction of technology (SCOT)

In science and technology studies, the SCOT approach opened up technological change to sociological analysis (for example, Pinch and Bijker, 1987; Bijker, 1995; Kline and Pinch, 1996). The SCOT approach shares with evolutionary economics the notion that technological change can be seen a process of trial and error, of variation and selection. But it differs from evolutionary economics in the sense that it argues that both variation and selection are fundamentally human processes. SCOT's evolutionary representation does not deal exclusively with artefacts, but consists of three layers: variation and selection of (i) problems, (ii) solutions and (iii) the resulting artefacts (Bijker, 1995: 51).

A central tenet in the SCOT approach is that social groups are relevant for understanding the development of technology (for example, users, producers, marketing departments, policy makers). An artefact may have different meanings for different social groups, for example, what they like about it, what they experience as problems, what they see as possible solutions. Relevant social groups may have different ideas about problems and solutions, and attribute different meanings to an artefact.[1] The artefact has 'interpretative flexibility'. The variety in definitions, meanings and specifications about an artefact does not endure, but is reduced through 'closure', which is primarily an intergroup process. Closure means that the interpretative flexibility of an artefact diminishes. Some problems are selected for further attention (setting of problem agendas). A variety of solutions is then generated. Some of these solutions are selected and yield new artefacts. Closure means that one interpretation of an artefact becomes dominant among the relevant social groups and others cease to exist. As part of the same movement, the dominant artefact will develop an increasing degree of stabilisation. This artefact then competes on the market with other artefacts.

In a later phase the concept of 'technological frame' was added to the SCOT approach (Bijker, 1995):

> A technological frame structures the interactions among the actors of a relevant social group. . . . Technological frames are located between actors, not in actors or above actors. A technological frame is built up when interaction 'around' an artefact begins. If interactions move members of an emerging relevant social group in the same direction, a technological frame will build up. Continuing interactions give rise to and are structured by a new technological frame. . . . A technological frame comprises all elements that influence the interactions within relevant social groups and lead to the attribution of meanings to technical artefacts. These elements include: goals, key problems, problem-solving strategies (heuristics), requirements to be met by problem solutions, current theories, tacit knowledge, testing procedures, and design methods and criteria. (p. 123)

In this way, action was conceptualised as taking place in a social and cognitive context. Artefacts also play a crucial role. Technological frames are not purely cognitive, but also comprise social and material elements. Actors who are members of different social groups will have different degrees of 'inclusion' in technological frames. The degree of inclusion indicates to what extent an actor's (inter)actions are structured by that technological frame.

A technological frame is located at the level of a relevant social group. Thus a technological frame needs to be sustained continuously by actions and interactions. The social construction of an artefact, the forming of a relevant social group and the emergence of a technological frame are linked processes (ibid.: 193). The existence of a technological frame can be used to explain the stability of an artefact and social group. There are also incentives for change, due to the different degrees of inclusion in a technological frame. The partial inclusion in different technological frames creates opportunities to work on novelties. The concept of a technological frame thus allows us to analyse both change and constancy in technological development.

An important insight from the SCOT approach is that technological development involves multiple social groups from the start, not only engineers, but also users, policy makers and so on. As these groups interact, a technological frame gradually develops and the form of technology stabilises (closure). These insights are an important building block for Chapter 3. But the SCOT approach says little about the wider diffusion of new technologies and the replacement of old ones.

Socio-cognitive approaches

The insight that shared cognitions and interpretations are important in technological change can also be found in other disciplines. The basic notion is that human beings have limited cognitive capacities to understand the world. Human beings are 'boundedly rational'. Social psychologists have shown that human beings use cognitive rules and interpretative schemas to understand the world, and assist decision making and (inter)action. Herbert Simon (1957) linked the limits of individual cognitive capacity to the nature of organisational structure. Organisational rules, procedures and cognitive routines not only support work and decision making, but also provide continuity and stability. Symbolic interactionism, phenomenology and neo-institutional sociology have argued that shared symbols and cognitions lead to shared constructions of reality and meaning (Goffmann, 1961; Berger and Luckman, 1967; Scott, 1995). Shared socio-cognitive institutions have coordinating power, and provide continuity and stability for action.

This socio-cognitive perspective does *not* mean that actors are passive 'cultural dopes', determined by a world of symbols, frames and rules. While the frames and rules are resources for action, actors are also active in modifying and changing them. Giddens (1984) coined the term 'duality of structure' to indicate that structures (such as rules and frames) are both the context and outcome of action. Cognitive structures can be changed on the basis of an evaluation of actions, that is, learning.

Socio-cognitive perspectives have also been applied to technological change (for example, Stankiewicz, 1992; Garud and Rappa, 1994). For instance, the SCOT approach is a socio-cognitive approach, with its emphasis on problem definitions, interpretative flexibility and technological frames. Other authors, including Garud and Rappa, highlight the co-evolution among beliefs, evaluation routines and artefacts. Beliefs indicate what is or is not feasible in R&D, for example, search heuristics or rules of thumb that researchers employ to address technological problems. Evaluation routines manifest themselves in a community of researchers, for example, testing routines and normative values that sustain and define the technology. Garud and Rappa argue that artefacts, beliefs and evaluation routine mutually influence one another as they develop over time. Researchers' beliefs influence the direction in which they work to improve artefacts. But not all beliefs have similar viability: 'The evaluation routines filter data in a way that influences whether or not researchers perceive information as useful' (ibid.: 347). Evaluation routines have a tendency to reinforce an established paradigm and preclude the emergence of others. The activities of engineers and researchers lead to artefacts, which are put in the market, where they encounter users: 'Physical artefacts set sense-making in motion as individuals interpret artefacts. Artefacts are cognitively worked upon by categorizing them with reference to existing beliefs' (p. 346). New artefacts may be interpreted using old cognitive categories. But as people learn about the new artefact, their beliefs and user preferences may change. This, in turn, may influence researchers' beliefs and evaluation routines, which then influences their activities and the direction of technological change. This leads to a 'cyclical dynamic in which technological paths are created' (p. 358). This conceptualisation has similarities to the socio-technical dynamic I described in Chapter 1 (Section 1.3).

A particular aspect of the socio-cognitive perspective has been highlighted in innovation studies and the history of technology, namely that new technologies are often initially interpreted using rules and categories associated with the old technology. This insight is an important building block for Chapter 3. Historians have given empirical examples of this dynamic:

> When drastically new technologies make their appearance, thinking about their eventual impact [that is, function], is severely handicapped by the tendency to think about them in terms of the old technology. It is difficult even to visualize the complete displacement of an old, long-dominant technology, let alone apprehend a new technology as an entire system. . . . The extent to which the old continues to dominate thinking about the new is nicely encapsulated in Thomas Edison's practice of regularly referring to his incandescent lamp as 'the burner'. Rather more seriously, in his work on an electric meter, a biographer reports, Edison for a long time attempted to develop a measure of electricity consumption in units of cubic feet. (Rosenberg, 1986: 24–5)

> Whenever a new technology is born, few see its ultimate place in society. The inventors of radio did not foresee its use for broadcasting entertainment, sports and news; they saw it as a telegraph without wires. The early builders of automobiles did not see an age of 'automobility'; they saw a 'horseless carriage'. Likewise, the computer's inventors perceived its role in society in terms of the functions it was specifically replacing in contemporary society. The predictions that they made about potential applications for the new invention had to come from the context of 'computing' that they knew. Though they recognized the electronic

computer's novelty, they did not see how it would permit operations fundamentally different from those performed by human computers. (Ceruzzi, 1986: 196)

Because the transistor was perceived as a replacement for a triode in a circuit . . . features peculiar to thermionic tubes were transferred to the transistor. Consequently, the idea of the integrated circuit developed slowly. The example of vacuum tube technology persuaded transistor manufacturers to treat semiconductors as if they were miniature, solid-state tubes. Borrowing from tube nomenclature and operation, the two transistor contact points and their electrical connections were respectively called *emitter* and *collector*, even though neither emitting nor collecting was taking place. (Basalla, 1988: 45–6)

When experience is limited, the customer's search for understanding is dominated by attempts to relate the new product to existing concepts. . . . In the early stages, the new product is defined largely in terms of the old; as learning occurs, it develops a meaning and definition of its own. . . . It was, thus, no accident that early customer decisions about automobiles were framed in terms of a choice between a 'horseless carriage' and a 'carriage with a horse'. (Clark, 1985: 245)

The reason for this conceptualisation with 'old' categories is that novelties do not emerge in a vacuum, but in an established context, surrounding an incumbent technology. Users have grown accustomed to the 'old' technology and producers have built up certain bodies of knowledge and cognitive routines. These established cognitive frames guide the development of new technologies. Established technologies may experience particular problems, engineers have particular competencies and problem definitions, users have particular preferences and so on. These problem definitions and preferences provide direction to innovative activities with new technologies:

[M]ost inventions have their origin in the attempt to solve very specific, even narrowly defined problems. In the early years, railroads were thought of merely as feeders into the existing canal system, to be constructed in places where the terrain had rendered canals inherently impractical. In precisely the same fashion, the radio was thought by its originators and proponents to have potential applications mainly where wire communication was impractical, for example, ships at sea and remote mountain locations. (Rosenberg, 1986: 24–5)

New technologies are often directed towards solving problems with the established technology. Radical inventions may first be interpreted as a small step, and only *ex post* be identified as major breakthroughs: 'Many major inventions were pre-adaptive in the sense that they were designed to solve a small local problem and mushroomed into something entirely different' (Mokyr, 1990: 286).When new technologies solve problems in the existing system, the relationship between old and new technologies is not necessarily competitive. Instead, there may be relationships of symbiosis and complementarity between old and new technologies.

The socio-cognitive conceptualisation highlights a particular dynamic in system innovations, namely, that the new grows out of the old, at least with regard to perceptions and interpretations. Novelties may be initially interpreted using existing rules. But the perceptions, rules and practices are gradually adapted and transformed on the basis of concrete experiences. Learning processes may lead to changes in concepts and rules:

> It is important to note that *concept development* is based in experience, and that it occurs through a sequence of interactions between the product and the user. Those interactions provide information about the new product's relationship to other products and to the customer's needs. . . . Grouping of this sort gives way to distinguishing as experience accumulates. . . . Experience with the automobile, discovery of its possibilities, led to the evolution of the concept from 'horseless carriage' to the 'automobile', with a set of subordinate concepts that gave further definition: speed, mobility, endurance, payload and so forth. *Conceptual evolution* is evident in words used to add further distinction to the varieties of automobile. One could not imagine widespread use of words like 'roadster' or 'touring car' or 'coupe' in the 1890s when the 'horseless carriage' was first introduced. These were *dimensions and categorizations that rested on years of customer experience*, on new habits in transportation and, of course, development of the product itself. (Clark, 1985: 243–5;) added italics)

This means that there may be aspects of transformation in system innovation, in which the new grows out of the old. This idea is also a building block for Chapter 3. New technologies may gradually be emancipated from their early context and grow in their own right. This is often accompanied by the emergence of specialised (technical) communities, which direct their inventive attention to improving the novelty. Such a new technical community may develop its own distinctive knowledge base, search heuristics, rules and so on. Van den Ende and Kemp (1999) give a specific example of this dynamic. In their analysis of the emergence of the digital computer they emphasised how the new comput*er* regime emerged in the context of existing comput*ing* regimes and only gradually grew into its own.

Domestication

Another sociological approach is domestication, which highlights the user side of (radical) innovation. Domestication approaches argue that new technologies are not only bought and adopted, but also have to be integrated into daily life and user practices. Consumers are embedded in a network of heterogeneous and interacting elements. Consumer choices should not be analysed from inherent user preferences, but these preferences should be understood from the embeddedness of consumers in a wider network of social and technical relations (Schwartz-Cowan, 1987). This opens up the 'black box' of diffusion. Adoption is no passive act, but itself requires adaptations and innovations.

This active dimension of adoption has been conceptualised under the heading of 'domestication'. Technologies have to be 'tamed' to fit in user routines and application contexts. New technologies have to be transformed from being unfamiliar and possibly threatening things to familiar objects embedded in the routines and practices of everyday life. Domestication involves (a) symbolic work, (b) practical work, in which users integrate the artefact in their user practices, and (c) cognitive work, which includes learning about the artefact (Lie and Sørensen, 1996). This means that consumption and adoption are themselves acts of innovation.

Users and other relevant actors from the selection environment (for example, policy makers, societal groups) play active roles in the domestication of new technologies. Impacts and effects are not simply *caused* by technologies, but *emerge* from learning and interaction processes. In his book *Electrifying America*, David Nye (1990), describes how electricity was gradually integrated in domains such as the factory, urban transportation, homes and rural areas. The transformations in these domains emerged gradually, as workers, managers, city authorities, housewives, farmers and so on learned about electricity and made it their own. This involved social struggles (for example, about legislation) and societal debates.

An important insight from domestication studies is that new technologies do not simply diffuse, but have to be integrated into user practices, regulation and wider socio-cultural meanings. A disadvantage of the focus on user practices is that technology may become a black box.

Social mechanisms

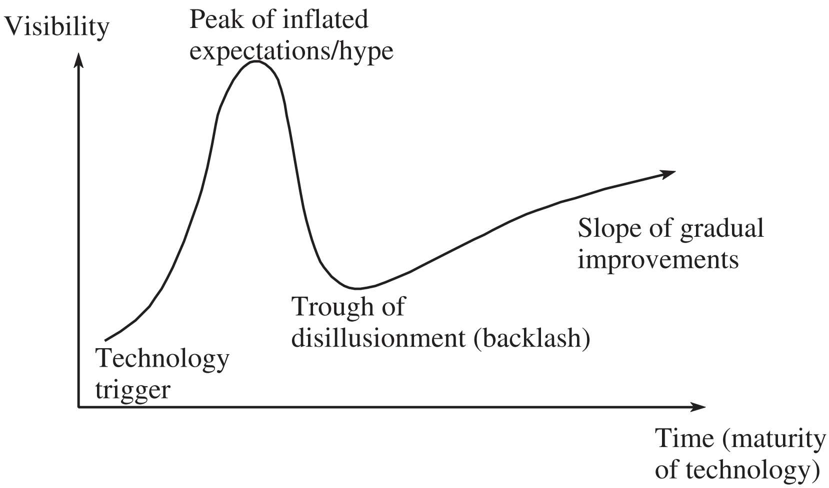
In sociological approaches the focus is on social groups and their interactions. These interactions may be characterised by social mechanisms, which influence the speed or direction of technological change. In the sociology of technology and business studies, several such mechanisms have been identified.

A first well-known mechanism is strategic games between firms. Firms do not operate in a vacuum, but position themselves in relation to each other. They play 'strategic games' through which they try to win market shares. Investment and adoption strategies of firms mutually influence each other. Inertia and acceleration are the outcome of two interacting considerations. The first is that the development of new technologies involves risk. Incumbent firms may alienate existing customers or lose sunk investments by developing new technologies. Furthermore, firms run financial risks when the development costs are high and the market uncertain. When the new technology fails, firms lose a lot of money. Other firms may adopt a wait-and-see attitude, letting the first firm carry the development costs. If the technology is successful, they may imitate or improve it with less cost and risk. The second consideration is that the development of new technologies involves opportunity, for example, 'first mover' advantages. The firm that first enters a new market may be able to command high market shares, enjoy publicity, boost their brand name, and establish linkages with customers. When the emergence of new technologies is characterised with much uncertainty, the first consideration plays a strong role. Incumbent firms may be trapped in a 'cartel of fear'. Firms watch one another, but none wants to take the first step. Even when they think that the new technology has potential, they are unwilling to bear the risk alone. The effect is inertia and slowing down of technical development. The inertia can be overcome, and may even result in accelerations when one player decides to adopt the second strategy and actually starts developing the new technology. When other firms fear that they may miss an emerging market, they suddenly find themselves caught up in an innovation race. Because of domino and bandwagon effects, technological developments may suddenly speed up.

A second mechanism is the importance of outsiders for the introduction of radically new technologies. Because incumbent actors have vested interests with the established technology, they tend to leave radically new technologies 'on the shelf'. If outsiders enter a particular sector, they may accelerate the market introduction of a new technology.

A third mechanism is that technologies may become part of 'hypes'. While a bandwagon effect is usually located in a sector, hype takes place in the wider society. It occurs when newspapers, consultants, the media and societal groups see a new technology as very promising (or threatening) and begin talking and writing about it. Great enthusiasm may be accompanied by much speculative investment in stock markets. An example is the Internet hype, which was thought to lead to a new kind of society. Third-generation mobile telephony (UMTS) was another hype. Both hypes were followed by a backlash, as promises did not materialise. Because telephone companies had paid large sums for UMTS licences, they now have great debts. Consultants at Gartner have summarised this dynamic using the hype cycle (see Figure 2.3), which consists of four phases:

1. *Technology trigger* A breakthrough, public demonstration, product launch or other event generates significant press and industry interest.
2. *Peak of inflated expectations* Overenthusiasm, unrealistic expectations and a flurry of well-publicised activity by technology leaders results in some successes, but more failures, because the technology is being pushed to its limits.



Source: Gartner.com (accessed 23 February 2004).

Figure 2.3 The hype cycle

3. *Trough of disillusionment* Because the technology has not lived up to its inflated expectations, it rapidly becomes unfashionable, and the press either abandons it for the next hot topic or emphasises its failure to meet expectations.
4. *Slope of gradual improvements* Focused experimentation and solid hard work by an increasingly diverse range of organisations leads to a true understanding of the technology's applicability, risks and benefits. Commercial off-the-shelf methodologies and tools become available to ease the development process and application integration. The real-world benefits are demonstrated and accepted. Tools and methodologies are increasingly stable as they enter their second and third generations.

A fourth mechanism is that the diffusion of new technologies may speed up or slow down when they become part of a social struggle. The introduction of the grain elevator in the Rotterdam port in the early twentieth century was accelerated when it became part of a struggle over power and authority in the port. Grain traders adopted grain elevators to reduce their dependence on port labourers, thus breaking their negotiation power (Van Driel and Schot, 2001: 312). The adoption of elevators was thus a move in a power game.

A fifth mechanism is the 'sailing ship effect'. Actors associated with an incumbent technology do not put as much effort in improving their technology as they could (for example, because they have a comfortable oligopoly or cartel). But when the established technology is challenged by a new one, the actors suddenly put a great deal of effort in making improvements. The improved performance of the established technology is a mechanism that can slow down diffusion of a new technology. The term 'sailing ship effect' was coined by Ward (1967), who referred to improvements in sailing ships in the 1860s and 1870s when they were challenged by steamships. Howells (2002) has recently cast doubt on the prevalence of the sailing ship effect. He argues that sailing ships and steamships were employed in such different market segments that they did not really compete. Improvements in sailing ships may have been the outcome of competition between sailing ship firms.

The social mechanisms described above show that particular kinds of interactions between social groups may lead to slowing down or acceleration of technological development. These mechanisms usually span shorter time spans (in the order of years). As such they could function as smaller building blocks in a wider perspective on technological transitions and system innovations. The mechanisms also provide a partial answer to the second research question. But in the literature identification of mechanisms has occurred in a rather haphazard way. More work can be done here, and steps in that direction will be taken in Chapter 7.

Socio-technical Approaches

In the socio-technical literature, the emphasis is on linkages between heterogeneous components. It is argued that new technologies are heterogeneous from the start, and develop by linking more elements together. Co-evolution is thus present from the start. The main dynamic is gradual shifts in networks and seamless webs. This literature tends to focus on the emergence of new technologies (invention) and their linking to markets, but says less about diffusion and the replacement of existing technologies. The large technical systems theory has been described in Section 2.2 because its subject was systems. However, the kind of analysis in LTS approaches is socio-technical. Another important socio-technical approach is actor-network theory.

Actor-network theory (ANT)

The perspective of socio-technical linkages is most consistently developed in actor-network theory (Latour, 1987, 1991; Callon, 1991; Callon et al., 1992; Law and Callon, 1992). ANT has been involved in foundational and ontological debates with a polemic tone (for example, about non-human actors). Because of my interest in dynamics, I shall avoid these debates. Furthermore, I shall play down the semiotic aspects (for example, translation, enrolment), because of their micro focus on actors.

The term 'actor-network theory' highlights the importance of linkages. The hyphen between 'actor' and 'network' means that there are no actors without networks. Actors are configured by their position in networks, by linkages to other elements. It is the embeddedness in networks that gives actors identity and provides them with aims, motives and resources. In its radical form this goes for both humans and artefacts. Artefacts do not 'work' unless they are placed in a wider configuration that works, and modern societies cannot exist without technologies.

New technologies emerge from the start as heterogeneous configurations. In the early phase of a new technology, the network consists of few elements and linkages. As the network is expanded and more elements are tied together, a technology 'becomes more real'. Reality, in ANT, is something that depends on the length and stability of the networks. To illustrate this perspective, Latour (1987) described how the diesel engine gradually emerged as Rudolf Diesel tied more elements together. In the early phase, the network consisted of no more than 'Carnot's thermodynamics plus a book plus a patent plus Lord Kelvin's encouraging comments' (p. 105). Diesel then interested the German MAN factory to develop the Diesel engine. Because MAN built steam engines, they had the relevant skills and tools for making pistons and valves. By expanding his network, Diesel imported resources, working towards the progressive realisation of the engine. The number of elements tied to his engine increased, as Diesel added 'MAN plus Krupp plus a few prototypes plus two engineers helping Diesel plus local know-how plus a few interested firms plus a new air injection system, and so on' (p. 105). Thus innovation is about the accumulation of elements and linking them together in a working configuration.

Actor-network theorists also understand diffusion as process of creating socio-technical linkages. Innovation often takes place in a particular local practice, for example, the MAN factory. Diffusion of an artefact is not simply placing it in the market. Instead, the diffusion across time and space needs to be accompanied by an expansion of linkages in which the artefact can function, for example, test apparatus, spare parts, maintenance networks and infrastructure. Without such a network the artefact cannot function. The diffusion process is no simple autonomous process with black boxes gliding effortlessly through space as a result of their own inner strength. Instead, 'the spread in space and time of black boxes is paid for by a fantastic increase in the number of elements to be tied together' (ibid.: 108). Simple economic models of diffusion neglect the 'work', which underlies diffusion: 'Thousands of people are at work, hundreds of thousands of new actors are mobilised' (p. 135). As linkages expand, a new technology becomes more stable, a black box, which can be taken for granted. All the required work and activities move to the background:

> There was a point, about 1914, when people could accept the Diesel engine not as a prototype, but as a copy, a black box. . . . The more automatic and the blacker the black box, the more it has to be *accompanied* by people. The black box stops pitifully when there is no salesperson, no repairer, no spare part. . . . The black box moves in space and becomes durable in time only through the actions of many people. . . . A continuous chain of people using, testing, and believing is necessary to maintain it in existence. . . . If you observe the new groups which are tied together, you will see how machines work and why facts are hard. (ibid.: 137–40; italics in original)

Actor-network theorists make interesting analyses of early phases of technical development, but pay less (fruitful) attention to later phases (for example, diffusion). Their emphasis on the accumulation of elements and linkages is an important building block for Chapter 3. But while the number of linkages can still be mapped in the early stage of innovations, this is more difficult for diffusion, when thousands of actors are involved.

Actor-network theorists aim to cancel the distinction between micro and macro levels, but do not entirely succeed. Once associations stabilise and become robust, they form the context for newly emerging technologies. With their micro focus on translation, alignment and so on, actor–network theorists fail to see these meso-level structures. But this does not mean that these structures have no effect. More attention should thus be given to interactions between dynamics at multiple levels.

New technologies do not emerge in an empty world, but in a world that is already made up of existing networks. In some places, ANT has paid attention to the interaction between established networks and new innovations. This leads to a particular pattern in system innovations, namely that new innovations link up in a technical sense with established networks to improve the overall functioning. Steam engines, for instance, did not immediately compete head on with water wheels. Instead, complex hybrid arrangements were used in the early stages. Early steam engines were commonly used to supplement the action of a water mill by pumping the water from the lower mill pond back to the upper pond, thus enabling it to run over the water wheel many times. Although (in retrospect) this was a cumbersome and inefficient arrangement, it was an essential expedient since the steam engine prior to the development of the *rotational* steam engine was really no more than a water pump (Rosenberg, 1976: 203). Another example of an early symbiotic relationship is that railroads did not immediately compete with canals and water transportation, but were used as feeders to them. The new technology (train) thus helped the old technology (boats) (ibid.: 197). The idea of technical linkages between old and new technologies is an important building block for Chapter 3.

ANT also highlights wider transformations in socio-technical configurations. If we introduce a new element (for example, new technology) into an existing network, it does not leave the network unaffected, but may trigger further changes and transformations. This is also a relevant building block for Chapter 3. Latour (1991) gives the example of the Kodak camera, in which the wet plates of the camera are replaced by dry collodion film. This was not merely a technical substitution, but triggered further changes:

> [C]apitalists are replaced by other capitalists, and above all, average consumers replaced professional-amateurs. . . . A successful innovation requires the simultaneously building of a new object (the Kodak camera) and of a new market (the mass-market). . . . The amateur market was explored, extracted, and constructed from heterogeneous social groups, which did not exist as such before Eastman. The new amateurs and the Eastman's camera co-produced each other. (pp. 113–17)

Technical substitutions can thus affect wider socio-technical configurations. This means that technology may act as *catalyst for social change* (Sørensen, 2002):

> When technology is perceived as a catalyst, it is seen as an agent that facilitates or makes possible a destabilization of a given social order by other actors and thus enables the opening of new options. Whether the social order is changed or not depends on the presence of actors and efficient strategies to make use of the destabilized situation. Thus, social change is neither made through new technologies nor through new social strategies or juxtapositions of structures, but rather through new socio-technical constellations. (p. 22)

Economic Approaches

Economic literature focuses on technologies and firms, which compete in markets. Price, performance, learning curves and market shares are important variables. The focus is often on diffusion and technical substitution. Simplistic assumptions tend to be made about the emergence of new technologies. Because of the focus on technologies in markets, wider aspects of co-evolution receive less attention.

Technology life-cycle approach

The technology management literature developed the technology life-cycle approach, which distinguishes three phases in technological development: (a) birth and childhood, (b) adolescence and (c) maturity. Typically a technology life cycle is described by indicators such as output volumes, market share or performance, and follows an S-shaped pattern. The technology life cycle is often linked to themes such as firm strategies, and market and industry structures.

Birth and childhood A new technology emerges as a variety of designs. The technology has low production volumes and market shares. Technical learning processes take place, aimed at product innovation. There is not only technological uncertainty, but also uncertainty about markets and functions. Many new firms try to gain a foothold in emerging markets, but there are also many failures and dropouts. The industry structure is fluid, consisting of many small firms, and high rates of entry and exit.

Adolescence/growth Initial diversity gives way to increasing standardisation, leading to a dominant design. Product innovations improve a technology's design features, although at a lower rate than in the childhood phase. Process innovations gain in importance to improve production economics (Utterback and Abernathy, 1975). The increasing rate of process innovation indicates that the product is being embedded in machines, factory layout, materials, supplies and so on. Improved product performance and decreasing costs lead to rapid market growth. The main strategy of firms is sales maximising. Firms compete more strongly, leading to shakeout, concentration and stabilisation in the industry structure.

Saturation/maturity Growth rates slow down as markets become saturated and technical improvements face diminishing returns. Further concentration in the industry leads to an oligopolistic industry structure. Competition is based on cost minimising, using economies of scale. Market saturation may lead to overcapacity in the industry, intensified competition and market volatility. Producers try to diversify to other products and markets. Product innovations are incremental and focus mainly on appearance and external design.

Economic path-dependence perspective

Arthur (1988) and David (1985) focused on competition between new technologies and argued that suboptimal technologies can win. In their path-dependency perspective, there are increasing returns to adoption (IRA), meaning that the more a particular technology is used, the greater its attractiveness relative to its competitors. Arthur (1988: 591) identified five sources of IRA: (a) learning by using: the more a technology is used, the more is learned about it, the more it is improved; (b) network externalities: the more a technology is used by other users, the larger the availability and variety of (related) products that come available and are adapted to the product use; (c) scale economies in production, allowing the price per unit to go down; (d) informational increasing returns: the more a technology is used, the more is known among users; and (e) technological interrelatedness: the more a technology is used, the more complementary technologies are developed. Once a particular technology is somewhat ahead of the other technology (because of contingent factors, incidents or social mechanisms), its lead will widen because of the IRA mechanisms.

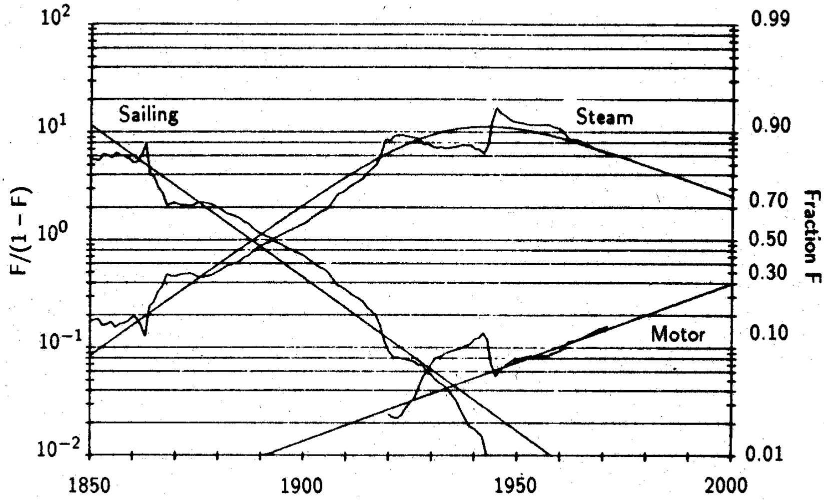
The relevance of these five IRA mechanisms for the diffusion of new technologies is a building block for Chapter 3. But these mechanisms say little about replacement of the existing technology. The path-dependency literature is unconcerned with the existence of prior technologies and existing contexts. It tends to focus more on the *emergence* of new technologies rather than *shifts* from one technology to another. Although path-dependency theories are useful to understand lock-in, they leave open the question of lock-out.

Technological substitution models

Grübler (1991, 1998) and Nakićenović (1986, 1991) understood long-term technological transitions as a replacement process. Their basic assumption is that the replacement of an old technology by a new one proceeds along the logistic substitution curve: $f/(1 - f) = e^{(a.t + b)}$, in which t is the independent variable representing some unit of time, a and b are constants, f is the fractional market share of the new competitor and $(1 - f)$ is the fractional market share of the old one. They tried to match historical data with these logistic curves. Figure 2.4, for instance, portrays the replacement of sailing ships by steamships in the United States and steamships by diesel motors. Logistic curve models are entirely descriptive, offering a phenomenological description of the substitution process, but no explicit explanation. Nevertheless, logistic curves suggest a particular kind of dynamic, namely a competitive struggle between old and new technologies.

Economic substitution approaches

The competitive dynamic is made more explicit in economic substitution approaches. Users are represented as having a set of user preferences. A new technology replaces an old one, if it is cheaper or performs better than the old one (a better price/performance ratio). Price and performance are not static properties of products, but develop dynamically over time. Hence, economic approaches have developed the concept of 'learning curves'. The basic idea is that the performance of technologies increases as firms gain experience with their functioning and production, that is, 'learning by doing' (Arrow, 1962). Unit costs decrease and performance increases along an exponential decay function. Because learning depends on the actual accumulation of experience, learning curves are generally described in the



Source: Grübler (1991: 166); courtesy of Elsevier Science.

Figure 2.4 Successive technological substitution of propulsive systems in the US merchant marine fleet, in fractional shares of gross tonnage, logit transformation

form of a power function, usually measured as cumulative output (Grübler, 1998: 82): $Y = a.\ X^{-b}$, where y is the cost or performance of the xth unit, a is the cost associated with the first unit, and b is a parameter measuring the cost reductions for each doubling of cumulative output (that is, learning rate). If the numbers on the axes are plotted as logarithmic scales, the learning curve becomes a straight line. The importance of learning curves in the economic substitution argument is that if the learning rate of new technologies is higher than that of established technologies then the former will eventually replace the latter.

There are several problems and weaknesses in the technological and economic substitution approaches. First, both logistic curves and economic substitution approaches suggest that old and new technologies have a relationship of competitive struggle. But in the subsections on sociological and socio-technical approaches (above) we have seen that this need not be so. Old and new technologies may also have symbiotic relationships, particularly in the early phase.

Second, neither substitution approach pays any attention to wider co-evolution processes in system innovations. Economic approaches take a limited cross-section of socio-technical systems. They look only at technologies and markets, and neglect other aspects, such as policies and regulations, user practices, infrastructures and symbolic meanings. Hence, they see diffusion merely as a market-based process, driven by price and performance. But this neglects all kinds of other elements and processes that influence diffusion. Rosenberg (1976) has given several reasons why a new technology may establish its advantages only slowly, including: (a) gradual technical improvements; (b) the development of technical skills among users; (c) the development of skills in machine-making; (d) technical complementarities; (e) improvements in 'old' techniques (the sailing ship effect); and (f) social, legal and institutional variables in the societal context. This means that we need a broader view than mere substitution approaches to understand diffusion.

Third, neither substitution approach is clear about the emergence of new technologies, which suddenly seem to emerge as a discontinuity. In some technology management literature, discontinuities are defined as 'relatively rare and tend to be driven by individual genius' (Tushman and Anderson, 1986: 440) or as 'stochastic breakthroughs' (Anderson and Tushman, 1990: 605). Such a conceptualisation of radical technical change is simplistic and mystifying. Perhaps if we look at successful outcomes and technical success, then a discontinuity can be found. The Wright brothers' aeroplane succeeded where others failed. But if we look at the underlying activities of invention (for example, articulation of problems in technical communities, gradual build-up of shared knowledge base) then invention is gradual and continuous. It is misleading to look only at great inventors, and present them as heroes or geniuses. This neglects the wider inventive context and how great inventors built upon earlier work by others.

Fourth, the demand side and user preferences are seen as fixed and static. Although user preferences may be assumed relatively stable in the short run, they definitely change over longer time periods. Furthermore, adoption is seen as a passive act, something that has been challenged by domestication studies, described in the subsection on sociological approaches (above).

Evolutionary economics

The focus in evolutionary economics (EE) is often primarily on firms, and secondarily on technology. The aim is not to explain technological change, but economic change. Technology is only relevant because of its influence on the firm's profitability. Furthermore, EE has more focus on incremental than on radical technical change. Hence, there are only a few contributions of interest to system innovation.

Building upon Joseph Schumpeter and the ideas of Herbert Simon and James March, Nelson and Winter (1982) and Dosi (1982) developed a new theory of the firm, based on bounded rationality and routines: 'The routinisation of activity in an organisation constitutes the most important form of storage of the organisation's specific operational knowledge. Organisations remember by doing' (Nelson and Winter, 1982: 99). The importance of routines is that 'the behaviour of the firm can be explained by the routines they employ. Knowledge of routines is the heart of understanding behaviour' (p. 128). Organisational and cognitive routines not only play a role in the *stability* of firms, but also with regard to invention and innovation. Dosi, Nelson and Winter looked closely at what engineers and designers actually do in firms, and found that their search activities were guided by cognitive heuristics or routines. Instead of searching in all directions, engineers and R&D managers typically expect to find better result in certain directions. Drawing on an analogy with Kuhnian paradigms in science, Dosi (1982) proposed the concept of 'technical paradigms' to understand the structured work of engineers:

> Technological paradigms are 'models' and 'patterns' for finding solutions to *selected* technological problems, based on *selected* principles derived from natural sciences and on *selected* material technologies. . . . A technological paradigm embodies strong prescriptions on the *directions* of technical change to pursue and those to neglect. (p. 152; italics in original)

Paradigms encompass exemplars and search heuristics. Because cognitive routines lead to patterned activities of search, they are the basis of path dependency and stability over time. In so far as firms differ in their organisational and cognitive routines, there is variety in their search directions and the resulting products. When different firms share particular routines, these routines make up a technological regime, something very similar to technical paradigms. Because search activities are focused in particular directions, they add up to technical trajectories on a sectoral level. Technological regimes are cognitive and relate to the technician's beliefs. Technological regimes create stability, because they provide a direction for incremental technical development.

The attention on stability at both firm and sector levels is a useful antidote to those approaches that too easily assume that new technologies simply replace old technologies (for example, substitution approaches). But Nelson, Winter and Dosi say little about how stable routines are overcome, how technological regimes come into being or how a change from one regime to another occurs.

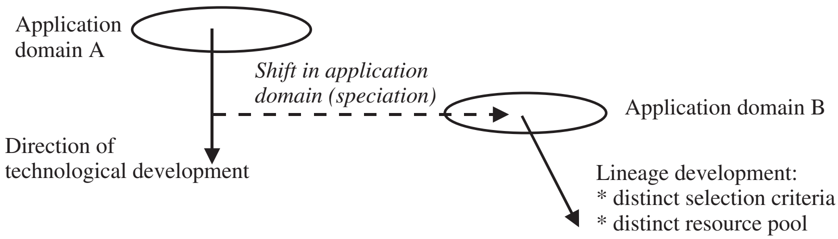
Some of these topics were recently addressed by Levinthal (1998), who argued that a major source of new technologies is speciation, that is, the application of existing technology to a new domain of application. Just as geographical separation may lead to a new biological species, a technological shift to a new market niche may trigger a new technological trajectory. Although the initial technological shift from one application domain to another may be small, the speciation event may trigger a new and divergent technological trajectory, if the selection criteria in the new application domain are sufficiently different. The *speed* of trajectories corresponds to the resource abundance of the market niche. The user preferences and selection criteria influence the *direction* of development (Figure 2.5). It should be noted, however, that the direction of technical change depends not only on selection criteria but also on shared cognitions in technical communities, for example, technical paradigms and regimes. Furthermore, Levinthal presents the 'shift in application' domain as too sudden, driven by visionary entrepreneurs. It is better to see it as a gradual articulation and learning process, involving multiple actors.

Levinthal highlights the importance of niches for the emergence of new technologies:

> the new technology finds an initial toehold for viability in the competitive ecology of alternative technologies in some specialized domain of application. . . . Often that initial niche are applications, such as the military, where distinctive performance on some dimension of functionality is highly valued and provides a basis for competitive viability. (p. 242)

This means that new technologies do not immediately compete with and substitute established technologies. This idea about the role of niches is an important building block for Chapter 3.

Niches are also important to understand diffusion. The critical factor for diffusion and substitution is whether the new technology remains in the initial, small niche, or is able to penetrate a broader set of niches. A similar notion was advanced by Schot (1998), who coined the term 'niche branching'. He distinguished two kinds of branching. The first kind refers to functional niches or application domains: 'Gas light was first applied in textile manufacturing, and was then used for street lighting and public



Source: Levinthal (1998: 223); courtesy of Oxford University Press.

Figure 2.5 Speciation in technological development spaces, such as theatres and cafés, and by wealthy individuals' (ibid.: 195). The second dimension of branching is a geographical expansion of markets. Gaslight, for instance, was first used in London, and then spread to other cities in England and to the Continent. The importance of niche branching is that momentum is generated. As a technology penetrates successive niches, it benefits from learning effects and scale advantages, improving its performance and driving down costs. As suppliers orient themselves to the new technology, more socio-technical linkages are created. Path dependencies and further irreversibilities occur for a variety of reasons: increasing returns, *de facto* standard setting, and sunk investments in competencies. The accumulation of learning processes and momentum makes the new technology more competitive so that it may eventually invade mainstream markets, possibly substituting the established technology.

2.4 CO-EVOLUTION

Co-evolution is emerging as an important concept in a range of disciplines, for example, evolutionary economics, innovation studies, industrial economics and long-wave theories. It has always been an important theme in science and technology studies, with its emphasis on seamless webs, emerging linkages between heterogeneous elements and co-construction. Different aspects of co-evolution have been dealt with, for example:

- co-evolution between technology and users (Clark, 1985; Leonard-Barton, 1988; Lundvall, 1988; Lie and Sørensen, 1996; Coombs et al., 2001; Oudshoorn and Pinch 2003; Schot and De la Bruheze, 2003);
- co-evolution between technology, industry structure and policy institutions (Carlsson and Stankiewicz, 1991; Nelson, 1994a,b, 1995; Rosenkopf and Tushman, 1994; Van de Ven and Garud, 1994; Lynn et al., 1996; Leydesdorff and Etzkowitz, 1998; Leydesdorff, 2000);
- co-evolution of science, technology and the market (Callon et al., 1992; Stankiewicz, 1992);
- co-evolution of science and technology (Layton, 1971, 1976; Kline and Rosenberg, 1986);
- co-evolution of technology and culture (Du Gay et al., 1997; Van Dijck, 1998);
- co-evolution of artefacts, beliefs (of designers) and evaluation routines (testing standards, equipment) (Garud and Rappa, 1994); and
- co-evolution of technology and society (Freeman and Soete, 1997; Rip and Kemp, 1998).

Although co-evolution has been studied with regard to two or three aspects, there is no literature that looks at co-evolution in socio-technical systems. A broader study of co-evolution is lacking. Co-evolution is often used as a reminder to disciplinary scholars that more aspects are important than they actually study. Co-evolution of technology and society, for instance, is an interesting phrase, but is mainly used as a flag to remind economists that there is more than 'meso' structures (for example, Freeman and Soete, 1997; Rip and Kemp, 1998).

In this section I shall organise different literatures by making an analytical distinction in different levels of co-evolution. The first level is that of newly *emerging* innovations. The emergence of *new* innovations can be analysed as a co-construction or alignment process, gradually linking heterogeneous elements together into a working configuration. Boundaries between the elements are created in the process. This approach is advocated in science and technology studies and innovation studies. The second level is that of *established* socio-technical configurations. The analysis assumes the existence of separate, but interdependent subsystems. Co-evolution can be conceptualised as relatively autonomous 'streams', which influence one another. This perspective is advocated in the triple helix approach. The third level is that of the co-evolution at the level of entire societies (macro level). Historians such as Fernand Braudel and recent long-wave theorists have tried to understand long-term developments as the result of pervasive technologies, demographics and cultural developments.

Co-evolution at the Level of Emerging Innovations

Many co-evolution literatures focus on the micro level of new, emerging technologies. The main dynamic is co-construction, that is, new social and technical elements shape each other. These elements are gradually created and linked together. During the period of flux, the boundaries among the elements are unclear and mixed by actors themselves. Stabilisation leads to the creation of boundaries among social, technical, political and economic elements. Shaping technology and building society are two sides of the same coin (Bijker and Law, 1992). Several literatures have highlighted aspects of this dynamic.

Science and technology studies (STS)

Co-evolution has always been an important theme in STS, with its emphasis on seamless webs, negotiation between multiple social groups, and emerging linkages between heterogeneous elements. The large technical systems theory, actor–network theory and social construction of technology approach describe how new technologies and user contexts are shaped simultaneously. LTS emphasises the work of system builders, who mould technologies, economics, regulations and user preferences into a configuration that works. ANT uses a more abstract approach to describe how elements are shaped and linked together (for example, translation, enrolment, alignment). Actor-network theorists argue that technology gradually becomes more 'real' as more elements are linked together, while it develops from an invention to a material prototype used in experiments, to a commercial artefact sold on the market. The later versions of the SCOT approach, another stream of STS, argue that cognitive, social and material elements are gradually woven into a technological frame as social groups and technical artefacts stabilise.

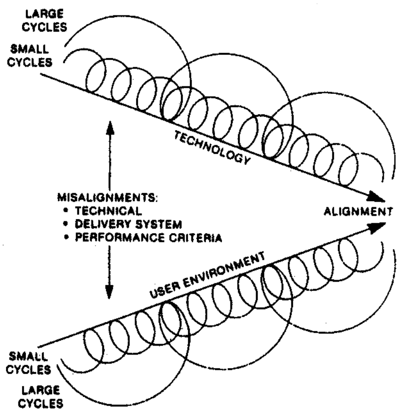
Co-construction of technology and users

Domestication studies and SCOT argue that users are not passive innovation consumers, but active agents, which provide crucial information for technological development. The importance of users has been highlighted in a recent book on the co-construction of users and technology, edited by Oudshoorn and Pinch (2003). Schot and De la Bruheze (2003) argue that the co-construction of products, consumption and consumers is mediated. Firms take the initiative to test user preferences (for example, test laboratories) and consumer representation groups are involved in design decisions.

Innovation studies have also emphasised co-evolution and the role of users in technological development. The involvement of users is crucial because they provide feedback about how a new technology matches their preferences and other aspects of their user context. Rosenberg (1982) introduced the term 'learning by using' to indicate the importance of concrete user experience in technological development. Lundvall (1988) described innovation as an interactive learning process, where user–producer interactions are crucial. Users may articulate unforeseen problems and propose innovative solutions or new functionalities. The productive character of user involvement explains why producers of new technologies who are closely tied in to users, have a greater innovation success rate (Von Hippel, 1988). The co-construction process between technology and user context can be represented as a gradual alignment of variation and selection environment (see Figure 2.6), requiring adjustments in both domains.

The co-construction of technology and markets

For radically new technologies the market may not exist, in which case new markets have to emerge with new products. Institutional economists and some business historians see markets as formed and constituted by rules. Firms and customers need rules to enforce reliability and stable playing fields. For new technologies, stable markets and 'rules of the game' do not



Source: Leonard-Barton (1988: 251); courtesy of Elsevier Science.

Figure 2.6 Co-evolution of technology and user environment

exist. These rules are constructed together with new technologies. In her recent book *Pirates, Prophets and Pioneers*, Deborah Spar (2001) analysed several technological breakthroughs (for example, international shipping, telegraph, radio, satellite television), and showed that the emergence of rules was important for the wide diffusion of new technologies. She concludes:

> In order for commerce to grow in any uncharted territory there need to be rules. Not regulation necessarily, or even governments, just rules. There need to be property rights, for example, and some sense of contracts. In higher technology areas there need to be rules for intellectual property (who owns the operating system? Under what terms?) and provisions for standardisation (how do different products work together? Which technical platform becomes the norm?). Without these rules, commerce may still emerge, but it will not flourish. (p.: xviii)

The co-construction of new technologies and markets has also been dealt with in innovation studies (for example, Coombs et al., 2001). Using a case study of gene therapy, Martin (2001) showed that the innovation was much more than just getting a technology to work. Potential applications and user groups also had to be constructed. Two potential user groups were doctors and drug companies, and attempts were made to interest both groups. Central to this process of enrolment was the organisation of experiments and clinical trials (to convince doctors of technical performance) and the formation of corporate partnerships with pharmaceutical companies. The use of some new gene therapy technologies involved the introduction of new forms of medical work (new work practices, roles, responsibilities), a new regulatory regime and a new commercial infrastructure. In sum, a new socio-technical network needed to be created. The issue of new markets and technologies was also addressed by Green (1992), using a case study about a biotechnology innovation called monoclonal antibodies (as a cure for cancer). He showed that the innovation was not merely technical, but involved the construction of markets through the articulation of relevant institutions (for example, information channels, property rights, distribution systems). To create markets, firms embarked on 'constituency building', involving regulatory agencies, standard-setting bodies, users and external sources of finance. The innovation failed because the elements were not successfully aligned.

In business and marketing studies, the co-evolution of technology and markets is approached from a firm perspective. How do firms know which users to target and how do they know what these users want? Some recent studies emphasise that the co-evolution of technology and markets is a highly uncertain process, marked by setbacks and surprises, and with no guarantee of success (for example, Leonard-Barton, 1995; Lynn et al., 1996). Traditional marketing instruments (for example, surveys, focus groups, mall studies) do not work if both the final shape of the product and the market are uncertain (Leonard-Barton, 1995). Hence, firms face much uncertainty in the case of radically new technologies. Despite the uncertainty, however, business and marketing studies indicate that there are certain patterns in successful management practices, namely market experimentation, 'probe and learning' (Lynn et al., 1996). Successful firms engaged in a series of market experiments in the early phases of technical development. Probing and learning was initially more important than immediate success:

> These companies developed their products by *probing initial markets* with early versions of the products, learning from the probes, and probing again. In effect, they ran *series of market experiments*, introducing prototypes into a variety of market segments. . . . The approach at work in these cases might best be described as *probing and learning*. . . . Probing with immature versions of the product only makes sense if it serves as a vehicle for *learning about the technology*, and whether and how it can be scaled up, *about the market* . . . and *about government regulations* and the need for regulatory approvals. . . . Probing and learning is an iterative process. The firms enter an initial market with an early version of the product, learn from the experience, modify the product and marketing approach based on what they learned, and then try again. Development of a discontinuous innovation becomes a process of successive approximation, probing and learning again and again. (ibid.: 15–19; italics added)

This management logic is more experimental than analytic. The market and technology for the radical innovation may look entirely different at the end of the experimentation process from how they did at the beginning.

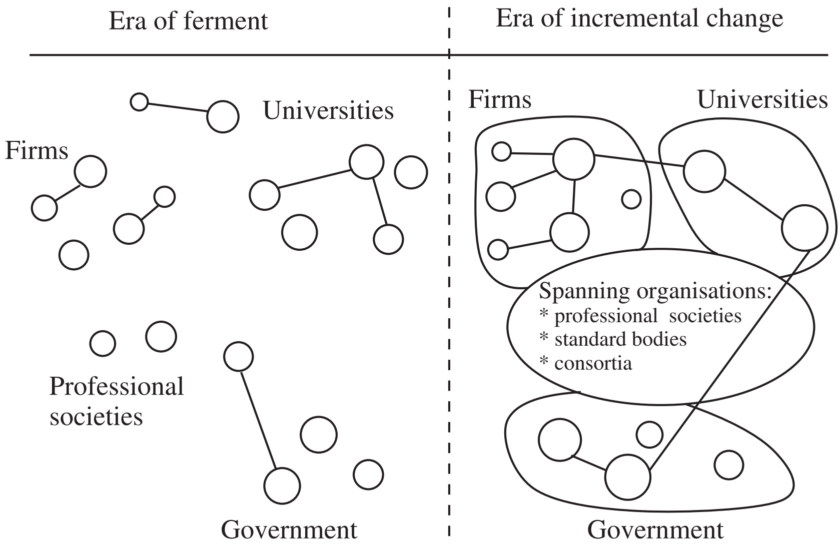
Although the final outcome is unknown, the probing and learning process is not blind. Instead, it is guided in each step by visions, expectations and ideas about possible uses. These visions and expectations are changed on the basis of the outcomes of learning processes. This emphasis on experiments, visions and learning is an important building block for Chapter 3.

Co-evolution of technology, industry structure and policy

In industrial economics the co-evolution of supply-side elements such as technology, industry structure and policy has been highlighted (for example, Tushman and Rosenkopf, 1992; Rosenkopf and Tushman, 1994). The interorganisational linkages in technological communities undergo extensive evolution in the shift from an era of ferment to an era of incremental change. In an era of ferment, when the technology is still in flux, the interorganisational network is fragmented. Few linkages exist and there are no spanning organisations, such as professional societies and standards bodies. Different coalitions of firms and universities work on different design variants. As a dominant design emerges, interorganisational linkages stabilise into a more organised network (Figure 2.7). Systematic interdependencies are built between organisations and industry-wide procedures, and traditions and problem-solving modes emerge (technical paradigms and regimes). The industry structure becomes more fixed, with a clear division of labour between participants. Spanning organisations (professional associations, standard bodies) are important for cognitive stabilisation, for example, articulation of problem agendas, standard setting. These spanning organisations also provide linkages between firms, universities and government. They engage in political lobbying to create favourable regulation for a new technology and public support programmes. Although Figure 2.7 has a bias towards interorganisational linkages, it nicely illustrates the dynamic of linking elements together into a stable configuration.

Co-evolution at the Level of Existing Configurations

There are also co-evolution approaches that focus on co-evolution at the level of existing socio-technical configurations. The focus is on interdependent, but relatively autonomous subsystems, which influence each other and co-evolve. Boundaries between the subsystems are the outcome of previous stabilisation processes. Because the boundaries have relative stability, the different subsystems have internal dynamics of their own. But the dynamics are also connected and related. This interrelatedness is represented in terms such as triple helix, techno-economic networks (Callon, 1991; Callon et al., 1992), techno-institutional complex (Unruh, 2000) and socio-technical system.



Source: Rosenkopf and Tushman (1994: 416); courtesy of Oxford University Press.

Figure 2.7 Interorganisational linkages during technological development

Triple helix

In the triple helix perspective the focus is on the ongoing co-evolution among three groups on the supply side of innovation: university, government and industry (Leydesdorff and Etzkowitz, 1998; Etzkowitz and Leydesdorff, 2000). The ongoing co-evolution is represented with three helices, interacting with one another through communication networks and linkages. An important consequence of co-evolution between dynamics in three helices is endless innovation. Perturbations in one sphere will trigger changes in the others:

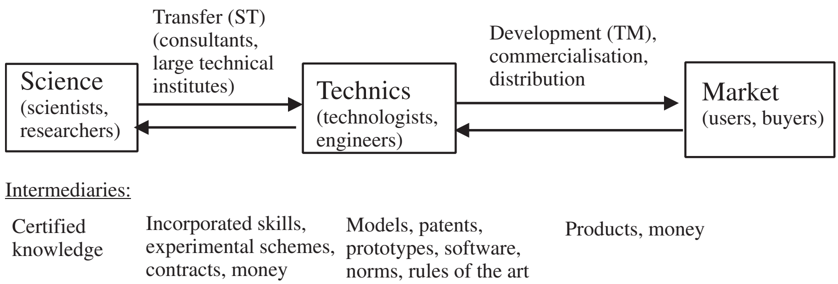
> In contrast to a double helix (or a coevolution between two dynamics), a triple helix, in which each strand may relate to the other two, is not expected to be stable. . . . The sources of innovation in a triple helix configuration are no longer synchronised a priori. They do not fit together in a pre-given order, but they generate puzzles for participants, analysts, and policymakers to solve. (Etzkowitz and Leydesdorff, 2000: 112)

During the co-evolution process tensions and a-synchronicities may emerge: 'Uncertainties in the relations between the helices open windows of potential innovation (and conflict) in (sub)systems that otherwise have to be reproduced' (Leydesdorff, 2000: 244). This is an important building block for Chapter 3. Socio-technical systems as described in Chapter 1 consist of more than three dimensions. Hence, we can expect tensions between ongoing dynamics in the constitutive elements. These tensions in existing configurations may create windows of opportunity for the emergence and diffusion of radical innovations. This means that wider circumstances (for example, processes in existing configurations) are important for radical innovations and that system innovations cannot be understood by looking only at the emergence of new innovations. It is also important to look at ongoing processes in socio-technical systems. Wider social and institutional changes may actually *precede* the emergence of new innovations.

Techno-economic networks

Co-evolution in existing configurations has also been studied in actor–network theory. As networks of heterogeneous linkages grow longer, they may develop into techno-economic networks with differentiated 'poles', representing a distribution of labour (Callon, 1991; Callon et al., 1992). A techno-economic network (TEN) is a 'coordinated set of heterogeneous actors, for example public laboratories, technical research centres, industrial firms, financial organisations, users, and public authorities, which participate collectively in the development and diffusion of innovations' (Callon et al., 1992: 220). In the original approach there are three main poles in a TEN: science, technology and market (see Figure 2.8). Different intermediaries circulate between the poles, for example, written documents (scientific articles, reports, patents, and so on), people and their skills, money (for example, contracts, loans, purchase) and technical objects (for example, prototypes, machines, products). The dynamics of scientific change, technological invention and the expression of demand are processes that are mixed up in one another. The TEN concept marks a kind of meta-coordination across the poles.

The network can be stable (cold, closed) or unstable (warm, fluid). In stable situations the shape of technology can be explained inductively, as a continuation of ongoing developments. Elements are strongly linked in convergent network. In unstable situations predictability is much less. Linkages between elements weaken, creating windows of opportunity for the introduction of new elements. The situation of the network thus determines whether a radical novelty is adopted or not. In cold situations the stability of the network provides a barrier for the novelty. As long as the existing network remains chained, convergent and long, there will be no opportunity for novelties to break through (Callon et al., 1992). If a network is 'heating up', windows of opportunity emerge for radical technologies to break through. This insight is a building block for Chapter 3.



Source: Callon et al. (1992: 222); courtesy of Elsevier Science.

Figure 2.8 The techno-economic network

Co-evolution at the Macro Level

At the macro level of entire societies, there are also some contributions that highlight co-evolutionary processes. One research approach is that of long-wave theories and economic history; another is general history. Some historians have tried to connect different strands of history in an overall synthesis.

Long-wave theories

Economic historians have found that prices and economic growth go up and down in cycles of about 50 years (Kondratieff cycles), signalling periods of boom and bust. The idea that radical innovations account for these cycles was first brought forward by Schumpeter, and later picked up by others such as Freeman and Perez (1988), Ayres (1989), Tylecote (1993) and Freeman and Louçã (2001).

To understand macro changes in the economy, Freeman and Perez (1988) coin the concept of 'techno-economic paradigm' (TEP) to indicate periods in which particular pervasive technologies, methods of production and economic structures reinforce and co-evolve with one another. They distinguish four historical successions in TEP, and claim that we are currently in the middle of the fifth cycle, related to information technology and biotechnology (see Table 2.3). In each new TEP, a particular input or set of inputs is the 'key factor'. This 'key factor' fulfils the following conditions: (a) clearly perceived low and falling relative cost; (b) apparently almost unlimited availability of supply over long periods; and (c) clear potential for the use of the new key factor in many products or processes.

Freeman and Perez (1988) aim to understand how changes occur from one TEP to another. One insight is that a new technology emerges in a

Table 2.3 Successive techno-economic paradigms

	Dominant technological system	Dominant methods and/or organisation	Key factor	Emerging technology
Industrial Revolution (1770–1830)	Water power, sail shipping, mechanisation in textiles	Manufacturing in factory	Cotton, pig iron	Mechanical techniques, steam engines, coal
Victorian Prosperity (1830–1880)	Steam power and iron (railway, steamships)	Centrally managed enterprise, joint-stock companies	Coal, transport	Steel, telegraph, electricity
Belle Époque (1880–1930)	Electricity and heavy engineering with steel	Standardised parts, large M-form corporations	Steel	Automobiles, aircraft, telecommunications, radio, oil
Golden Age of growth, consumer society (1930–1980)	Automobiles, aircraft oil and petrochemicals, synthetic materials (plastic)	Fordism/Taylorism, mass production	Energy (oil)	Computers, microelectronics, radar, drugs, nuclear power, numerically controlled machine tools
ICT (1980–?)	Computers, software, telecommunications, robotics, satellites	Quality control, globalised enterprises, de-centralised management, just-in-time management	Information (chips and microelectronics)	Biotechnology products, space activities, fine chemicals, artificial intelligence

Source: Freeman and Perez (1988: 50–57).

world that is still dominated by the old paradigm. The new technology demonstrates its advantages at first in one or a few sectors. This is similar to the point about niches made by Levinthal (1998) and Schot (1998). Old and new technologies need not be competitive from the start, because 'at first these innovations may appear as means for overcoming the specific bottlenecks of the old technologies' (Freeman and Perez, 1988: 48). The emergence of radical innovations is explained as a reaction to problems in the existing TEP, in particular perceived limits to growth, diminishing returns in productivity and a weakening of the key factor. This means that new technologies can remain in particular sectors, as long as the existing TEP is stable and not in crisis. In fact, the new technologies may not fit the institutional and social framework: 'Initially there will be a degree of mismatch between the techno-economic subsystem and the old socio-institutional framework' (p. 59).

Breakthrough of the new technology depends on processes in the existing TEP. As the key factor of the old paradigm runs into problems, the new technology acquires dynamics of its own. One aspect of these dynamics is the importance of cross-sectoral clustering between technologies (Ayres, 1989). The combination of steel, gasoline and the internal combustion engine was crucial for production and use of the automobile.

A third point is the occurrence of wider social changes to overcome the initial mismatch. The diffusion of the new technology is accompanied by a wide range of co-evolving changes in: (a) organisational forms in the firm and at plant level; (b) new skill profile in the labour force; (c) new product mix; (d) wave of infrastructural investment; (e) tendency for new innovator–entrepreneur-type small firms to enter expanding branches of the economy; (f) new pattern of consumption of goods and services; and (g) new types of distribution and consumer behaviour (Freeman and Perez, 1988: 59). Diffusion thus involves changes on many aspects.

These three points form relevant building blocks for understanding system innovations. But their analysis also has some weaknesses. Because of the macro focus, no detailed attention is paid to what engineers really do and think. Long-wave theories suffer from deterministic overtones, because there is no account of how new technologies emerge. The idea seems to be that they simply emerge from 'key factors', limits and bottlenecks in the existing TEP. This is too simple, because it ignores the variety of technologies available to solve bottlenecks. It thus needs to be explained why certain technologies win and others lose. In macro theories, such as the long-wave perspective, it is important to keep visible the micro level and the perceptions and activities of actors involved in developing new technologies. Macro aspects should be combined with micro processes in a coherent perspective. This means that we need a multi-level perspective.

In long-wave theories there is also the danger of technological determinism. Freeman and Perez's suggested causality is that techno-economic forces do the initial *acting* and the socio-institutional framework does the eventual *reacting*. Insufficient room is made for the less tangible social innovations, which co-exist alongside the harder technical ones. In the subsection on co-evolution at the level of existing configurations (above), it was argued that wider social and institutional changes may actually *precede* the emergence of new innovations. Bruland and Smith (2000) also make this point in a case study of steam power:

> Steam power appears to be not a technological driving force in the early nineteenth-century British economy, but rather a component of a much wider change, involving incremental technological shifts, and social and organisational changes. . . . All of this fits into a wider process of social change which comprises industrialisation: the emergence of finance sources, of disciplined work forces, of distribution systems for products, and so on. These things, of course, do not just emerge: they were created via long-term processes of social and institutional change. These changes were not simply effects of some prior process of radical technological breakthrough; in fact it would be more plausible to argue that *they were the preconditions for technological change, rather than its effect*. (p. 18; italics added)

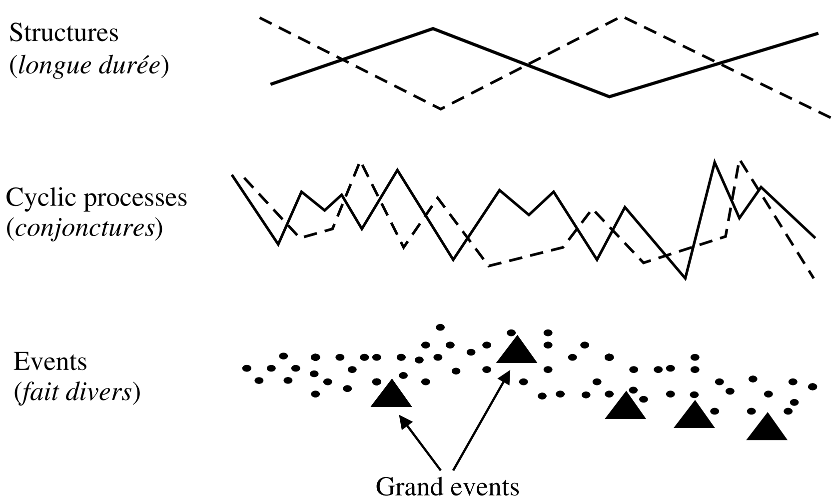
As a defence to the charge of technological determinism, Freeman and Louçă (2001) made further refinements to the long-wave theory. They distinguish five 'subsystems' in society: science, technology, economy, politics and culture, which are all equally important. Because each of these subsystems has its own distinctive features and its own 'selection' environment, they have relative autonomy. On the other hand, the subsystems also interact with one another, because activities of social groups are coordinated and aligned to one another. Long waves can be understood as co-evolutionary processes with periods of alignment and periods of disalignment between the five subsystems. The social subsystems generate a large number of irregular fluctuations and variations, caused by specific subsystem cycles (for example, political and business cycles, technological trajectories, cultural movements, life cycle of industries). These fluctuations are usually dampened by the linkages with other subsystems, which provide coordination. At times, however, the fluctuations result in major maladjustments and a lack of synchronicities. These tensions between subsystems are the underlying cause of periods of economic crises, and may be overcome through innovations in the subsystems and gradual realignment. Renewed positive congruence and interaction between the subsystems provides fertile soil for economic growth. Thus, 'it is essential to study both the relatively independent development of each stream of history and their interdependencies, their loss of integration, and their reintegration' (ibid.: 127). This notion of understanding major changes as dealignment and realignment processes between different elements is a useful building block for Chapter 3.

General history

Some general historians have tried to write synthesising works to weave together different strands of historical processes, for example, Johan Huizinga, Jan Romein, Manuel Castells and Fernand Braudel. The French historian Braudel is particularly interesting, because he developed a conceptual framework, which distinguishes different levels of historical time. Braudel has gained worldwide recognition with, at least, two works, *The Mediterranean and the Mediterranean World in the Age of Philip II* (Braudel, 1976) and *Civilization and Capitalism, 15th–18th century*, which encompasses three volumes *The Structures of Everyday Life, The Wheels of Commerce* and *The Perspective of the World* (Braudel, 1985a,b,c). Braudel refuses to be constrained by arbitrary disciplinary boundaries and connects different dimensions in his work. He writes about demographics, which he sees as fundamental to understanding history. Closely related to demographics is food provision, in the form of the three great cereal crops, wheat, rice and maise. Food provision, in turn, is related to environmental conditions such as droughts, floods, rainfall patterns and temperature, which influence food prices and trading patterns. Technological innovations are also important, for example, regarding energy sources, metallurgy, transportation, gunpowder, printing and sea navigation. Culture is also important, for example, in the form of fashions, table manners, and the use of luxuries such as salt, meat and spices. In his stories, Braudel pays much attention to the structures of everyday life; wars, treaties, kings and popes feature only incidentally. Their actions are events against the backdrop of wider social, cultural, economic, technological and environmental conditions. Nevertheless, he is able to combine detail with the broad picture, thus preventing any structural determinism, where macro conditions determine micro activities. Braudel's work is an attempt at co-evolutionary synthesis rather than at a summary of historical developments.

At the heart of Braudel's synthesis is a three-level hierarchy that distinguishes different speeds of historical developments. These levels range from superficial events and actions to 'deep history'. The macro level is formed by deep structures with rhythms of 50–100 years, forming the *longue durée* (Braudel, 1958). This level refers to aspects such as geographical landscapes, which influence communication and trading patterns (for example, the shape of the Mediterranean Sea, mountain ranges, rivers), demographics and environmental conditions (soil, climate, rainfall). These structures cannot be changed at will by human actors and provide the backdrop for action. The meso level refers to cyclic processes and gradual changes in domains such as agriculture, transport, the economy, military logistics, politics, cultural values, communication and trading patterns. The time-frame is that of decades. To understand dynamics at this level, social history is useful, in particular for looking at (changes in) interactions between social groups such as large landowners, the nobility, the bourgeoisie, urban craftsmen, peasants, traders, military leaders and city governments. The interaction between these social groups results in up- and downswings, as the balance of power and alliances change in one direction or another. Although cyclic processes are the outcome of (inter)actions of social groups, they cannot be influenced by individual human beings. The micro level is that of political, diplomatic and military events with a time length of months and years. This is the stuff of traditional historical accounts, which look at triumphs and failures of 'big men' (kings, military leaders). Some of them are large events, for instance, the peace of Cateu-Cambrésis, the Turkish siege of Malta, the Holy League and an endless procession of wars, battles and peace treaties. A sequence of events may go in particular directions, creating trajectories in history. But Braudel sees this level as superficial history, which cannot explain the whole of history. Figure 2.9 schematically represents Braudel's three layers of historical processes.

This perspective of three levels is a crucial building block for Chapter 3. But in Braudel's work there is a structuralist bias. He argues mainly for a



Source: Bertels (1973: 123).

Figure 2.9 Different historical time developments top-down explanation, where processes need to be understood from the structures in which they are embedded. To understand system innovations, I shall argue that this needs to be complemented with a bottom-up perspective, that is, changes in structures need to be understood from their constitutive processes.

NOTE

1. In their empirical analysis of the bicycle in the second half of the nineteenth century, Pinch and Bijker (1987) showed that the high-wheeled, 'Ordinary' bicycle had different meanings for different groups. For young, upper-class men, the Ordinary was a means to show off their athletic skills in parks. The fact that the bicycle was difficult to mount and risky to ride had a positive meaning, because it emphasised their skills. For them, the Ordinary was a machobicycle. For women and middle-aged men, however, the Ordinary was an unsafe bicycle, because of the risk of falling down and breaking a leg. One artefact thus had different meanings.

3. Conceptual perspective on system innovations and technological transitions

3.1 INTRODUCTION

This chapter develops a conceptual perspective on system innovations and technological transitions. This perspective builds on the insight from complex systems theory, general history and long-wave theory that major changes come about because processes at multiple levels link up and influence one another. Hence, the perspective is multi-level. I shall build upon an existing multi-level perspective on technological change, described in Section 3.2. Because this perspective has some weaknesses for understanding system innovations, I shall make several conceptual additions in Section 3.3, using building blocks from Chapter 2. The improved multi-level perspective is not a formalised model, but an 'appreciative theory' (Nelson and Winter, 1982); it addresses the first research question. The research question about patterns is addressed in Section 3.4. Three patterns will be suggested: (a) fit–stretch pattern in the co-evolution of form and function, (b) co-evolution of technologies and (c) diffusion as a trajectory of niche accumulation. The research question about mechanisms is not addressed in this chapter. It will be taken up in Chapter 7, building upon findings from the case studies.

3.2 THE MULTI-LEVEL PERSPECTIVE

The multi-level perspective (MLP) was originally developed to understand regime shifts (Kemp, 1994; Schot et al., 1994; Kemp et al., 1998; Rip and Kemp, 1998;Van den Ende and Kemp, 1999; Rip, 2000; Kemp et al., 2001). To that end, three levels were distinguished: technological niches, technological regimes and socio-technical landscape. The conceptualisation of dynamics at these levels aims to combine insights from evolutionary economics, innovation studies and science and technology studies.

The lack of integration of different perspectives has been signalled as a general problem, see for instance, Cronberg and Sørensen (1995: 20) and Williams and Edge (1996). In his assessment of technology studies (in the Netherlands), Van Lente (1995) arrived at the following general conclusion:

> [A] second weak point: the relative lack of cumulation of insights in spite of important opportunities for this. . . . There are a lot of important findings, but there is relatively little effort to collect these findings and prepare integrative pieces. . . . As a consequence, the opportunities for systematical theory development are not exploited satisfactorily. . . . Thus far, many activities in technology studies found their roots in the opposition against the dominant view of technology as a deterministic, yet exogenous (autonomous) factor. While the dominant view is still powerful, the 'social shaping' view is more and more accepted. This success implies that a next step of technology studies is timely: from contesting the dominant view of technology towards a further articulation of a programme, based upon the many important outcomes and the possibilities for further connections with established disciplines. (pp. 318–19)

Many authors, calling for an integration of perspectives, see a combination of evolutionary economics, and STS as promising (for example, Coombs et al., 1992; MacKenzie, 1992; Williams and Edge, 1996). For instance Weber (1997) thinks that:

> A major convergence can be identified between evolutionary economics and the sociology of technology. Although they have very different roots, the basic understanding of the process of technological change is quite similar, and – even more important – sufficiently open to introduce elements of the other perspective. . . . What is still missing is the actual integration in a single framework which would allow to investigate different cases from a wider perspective, and to bridge explicitly between economics and sociology with regard to technology studies. (p. 83)

In sum, a combination of evolutionary economics, STS and innovation studies seems fruitful. I shall not pursue this as a general question, but as a way to create a conceptual perspective on system innovation and technological transition. To that end, I shall describe the three levels of the multilevel perspective and their interaction. These levels aim to take into account the insight that socio-technical and co-evolutionary dynamics are played out at different levels, and it also builds on Braudel's notion of processes with different time-scales at different levels (see Chapter 2, Section 2.4).

Technological Regime

The concept of 'technological regime' stems from evolutionary economics, where it was coined by Nelson and Winter (1982) to explain the occurrence of technical trajectories. Technological regimes refer to cognitive routines that are shared by engineers and designers in different companies. Search heuristics, in particular, are an important kind of routine, because they influence the direction of innovation. This innovation is usually of an incremental kind, leading to optimisation of existing technologies. Rip and Kemp (1998) widened the definition of technological regime to make it more sociological:

> A technological regime is the rule-set or grammar embedded in a complex of engineering practices, production process technologies, product characteristics, skills and procedures, ways of handling relevant artefacts and persons, ways of defining problems; all of them embedded in institutions and infrastructures. (p. 340)

The sociological concept of rules is wider than cognitive routines. In Chapter 1, formal rules (for example, laws, regulations, standards, incentive structures) and normative rules (for example, norms, role expectations, behavioural values) were distinguished from cognitive rules. Formal and normative rules are also important in stabilising existing technologies and influencing technical trajectories. Another aspect of Rip and Kemp's definition is that rules are embedded not only in the minds of engineers, but also more widely in the knowledge base, engineering practices, corporate governance structures, manufacturing processes and product characteristics. These wider embeddings make rules harder to change.

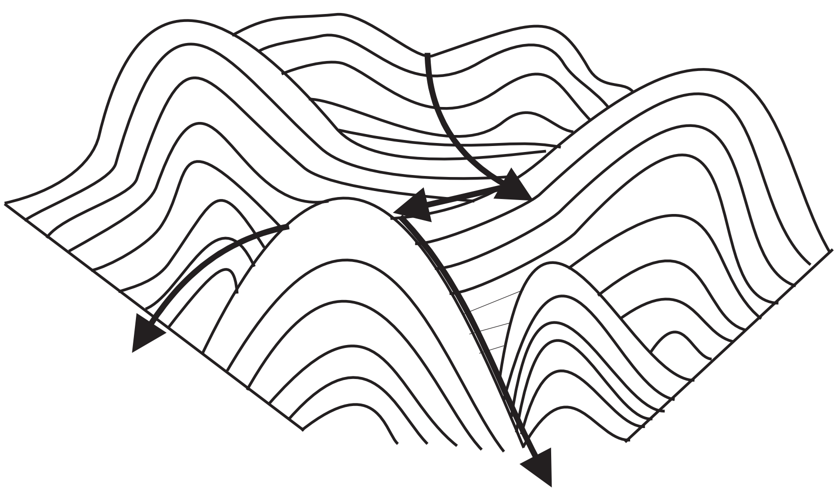
To avoid confusion about rules, let me emphasise that the relationship between actors and rules should *not* be understood as passive rule following. Rules are a resource for action because they provide guidance. But the application and interpretation of rules is not passive; instead it requires action and creativity. Rules may need to be modified to meet the contingencies of real-life local practices. While rules are a resource, they are also a constraint, making deviating actions more difficult. Rules thus have enabling and constraining effects. In Chapter 1, it was shown how rules and action have a recursive relationship. Existing rules provide a structure for (inter)action, but the effects of the interactions may change the rules. Rules are both the medium and outcome of action (Giddens, 1984). They may change gradually over time, during cycles of action and structuration.[1]

The importance of rules is that they provide coordination for human action and (inter)action. If rules are shared in social groups or communities, the activities go in the same direction, resulting in stability and coordination. This stability is not automatic. Instead, it is 'dynamic stability' in the sense that even stability requires 'work', for example, reproduction of rules or 'repair work'. Technological regimes account for the *stability* of technological development. This does not mean that technological regimes are static. Rather, they are dynamically stable, meaning that incremental innovation occurs along technical trajectories to improve the dominant design, leaving the basic design rules intact. Although the different rules in a regime are aligned, there may be times when there are tensions and misalignments. Hence, it is better to see technological regimes as a semi-coherent set of rules.

The Socio-technical Landscape

Technological trajectories are situated in a 'socio-technical landscape', consisting of a set of deep structural trends external to the regime. The analytical importance of 'landscapes' is that they form an external structure or context for interactions of actors. While regimes refer to rules that enable and constrain activities within communities, the 'socio-technical landscape' refers to a wider technology-*external* context, which cannot be changed at will by regime actors. The context of landscape is even harder to influence than that of regimes. In that sense the landscape is similar to Braudel's structures or *longue durée* (see Chapter 2, Section 2.4). With regard to technological trajectories, the landscape can be understood as providing a 'gradient' or backdrop for development (Figure 3.1). This means that the landscape makes it easier for technical trajectories to go in certain directions rather than others.

The content of the socio-technical landscape is rather heterogeneous and may include aspects such economic growth, emigration, broad political



Source: Sahal (1985: 79); courtesy of Elsevier Science.

Figure 3.1 Landscape as context for technological trajectories coalitions, cultural and normative values, environmental problems, resource scarcities, oil prices and wars. But the landscape also includes the large-scale material context of society, for example, the material and spatial arrangements of cities, factories, highways and electricity infrastructures (Rip and Kemp, 1998). In fact, the 'landscape' metaphor was chosen because of this literal connotation of material 'hardness'.

There are two kinds of changes in landscapes. The first landscape change refers to relatively slow changes, for example, cultural and demographic changes, and changes in political cultures and ideologies. This is in line with Braudel's *longue durée*. The second landscape change refers to relatively rapid developments, for example, war, oil prices and economic depression. This is somewhat at odds with the *longue durée*, but it is necessary to account for external shocks.

Niches

Radically new technologies usually emerge as 'hopeful monstrosities' (Mokyr, 1990). They are 'hopeful' because they can perform a particular function, but they are 'monstrous' because their performance characteristics are low. Most inventions are initially relatively crude, have low technical performance, and are cumbersome and expensive.[2] Because such innovations cannot survive in normal, mainstream markets, they need some protection. This protection is provided by niches, which are actively constructed by incubators or product champions. These actors are able to rouse the interest of other actors (for example, policy makers, users, suppliers) to mobilise resources for further development work of the innovation, by making promises about possible functionalities and market applications (Van Lente, 1993; Van Lente and Rip, 1998). Their activities lead to a support network of actors, who are willing to put time and money into nurturing and developing the innovation. Because niches offer some protection, they act as 'incubation rooms' for radical innovations.

Niche developments occur in two (partly overlapping) forms: technological niches and market niches. In technological niches, protection is provided in the form of public subsidies or strategic investments by firms (for example, Schot et al., 1994). In market niches, the protection comes from special-purpose performance requirements for special applications, and regular market transactions provide the resources to sustain the niche.

Another difference between technological and market niches is the degree of stabilisation. In technological niches rules are still unstable. There is uncertainty about design rules (for example, dominant design, search heuristics, exemplars), it is unclear who the users are and what their precise preferences are; there is neither a stable network of producers, nor any market institutions (for example, dealer network, property rights, formal rules). The social network, which carries the technology, is precarious and may experience large shifts in its structure and composition. Learning and articulation processes are important to stabilise and refine the rules. This refers to technical learning, but also to the articulation of user preferences or roles in production networks. These learning processes cost money and require effort. Niche actors are willing to sponsor these activities, because they have high expectations for the future, or because government subsidies lower the financial risks. Financial transactions occur in personalised networks, and involve extensive information costs.

In market niches there is more stabilisation. The technology has become stable enough that some producers go beyond the prototype stage and offer it for sale. There is also stabilisation in the sense that basic market institutions develop (for example, dealers who have brochures with price–performance statements about products). It has also become clearer what the users want from the technology. Specialised market niches may provide new technologies with protection, if the selection criteria are very different from the regime. An example is the Army, for whom high-performance requirements are often more important than cost. Hence, the armed forces have stimulated many radical innovations in their early phases (for example, digital computers, jet engines, radar). In market niches, learning processes are still important. Technologies can be improved, new user groups explored and regulations adapted.

The role of niches in the emergence of radical innovations has been studied under the heading of strategic niche management, SNM (Schot et al., 1994; Kemp et al., 1998; Schot, 1998; Hoogma, 2000; Kemp et al., 2001; Hoogma et al., 2002). SNM scholars argue that three kinds of processes are important for the construction and development of niches. The first is the building of a social support network to nurture novelties. Second are learning processes to stimulate the price/performance ratio of new technologies and their alignment in broader socio-technical systems. This includes learning by doing (Arrow, 1962), learning by using (Rosenberg, 1976; Von Hippel, 1988) and learning by interacting (Lundvall, 1988). Learning processes take place on multiple dimensions. On the supply side, these dimensions involve technology, production network and policies. On the user side, they involve user preferences, regulation, infrastructure, policies, symbolic meaning and maintenance networks. An important goal is the alignment between these dimensions, that is, co-evolution on a micro level (see Chapter 2, Section 2.4). The third process is the articulation and adjustment of expectations and visions. Expectations fulfil two functions: (a) they give direction to the learning processes, and (b) they are used to attract attention and enrol more actors to widen the social network.

If these processes reinforce one another, the niche will gradually expand and become more stable. Positive learning experiences may strengthen the expectation, which, in turn, may attract more actors to support the niche. The expanding social basis makes the niche more stable. But the gradual alignment of socio-technical aspects also makes it more stable, for example, the alignment of user preferences and technical characteristics, the emergence of markets and market institutions, and supporting legislation. The emerging stability will be represented by more stability in the rule-set, which guides the actions of actors. On the other hand, if the three internal niche processes do not reinforce one another, the niche may fall apart. When learning processes and technical performance do not live up to expectations, actors may leave the support network.

The Sociological Characteristics of the Three Levels

The multi-level framework is based on sociological concepts, with a focus on activities of people and the rules, which provide the context for action and interpretation. On the one hand people act on the basis of rules, but on the other people learn from their actions and may change the rules. The (socio)logic of the three levels is that they provide different kinds of structuration of activities in local practices.

The rules in technological niches are vague and imprecise, for example, diffuse promises about potential uses (Van Lente, 1993) and different ideas about best search directions. There are no specific rules that guide activities. The activities of actors in a niche go in many directions. There is experimentation and uncertainty. The niche rules provide only loose structuration. The rules have no strong coordinating 'power' on actors.

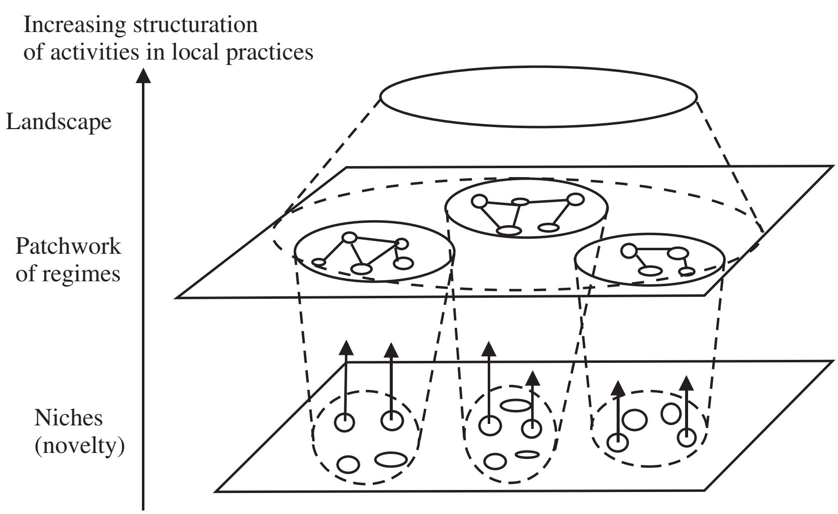
In technological regimes, structuration of activities in local practices is much stronger than in technological niches. A reversal has occurred. While actors put in a lot of effort to articulate and uphold rules in niches, the rules have become 'stronger' in technological regimes. They have strong coordinating effects on the activities of actors in local practices. The rules do not determine actions, but strongly guide them. It is possible to deviate, but this is difficult, and takes a lot of effort. Social conventions and rules in (technical) communities are not easy to change.

Socio-technical landscapes provide even stronger structuration of local activities. Material environments and widely shared cultural beliefs, symbols and values are hard to deviate from. They form gradients for action. Landscapes are more difficult to change than regimes.

Building on insights from hierarchy theory and complex systems theory (see Chapter 2, Section 2.2), I propose that the relationship between the three levels of the multi-level perspective can be understood as a nested hierarchy. The nested character means that regimes are embedded within landscapes and niches within regimes. In Figure 3.2, I have indicated this embeddedness with dotted lines. Multiple regimes are embedded in the same landscape, and multiple niches in a particular regime. Novelties at the niche level are indicated with small circles and arrows because they are often geared to the problems of existing regimes. Novelties are intended to be used in the regime, but they may face a mismatch with the existing regime, which hinders their breakthrough. Furthermore, the stability of the regime makes it difficult for them to break out of the niches. This stability emerges from the interlinkages between heterogeneous elements, also represented in Figure 3.2. This representation is schematic and stylised. In reality, the boundaries between different regimes are less clear, there may be overlaps, and boundaries may shift over time – niches may not fall neatly in one regime, but emerge at the boundaries between regimes; or they may begin in the context of one regime, but end up in another.

Dynamic Interactions

The MLP has been used to understand the emergence and diffusion of new technologies. An important point of the MLP is the dynamic interaction

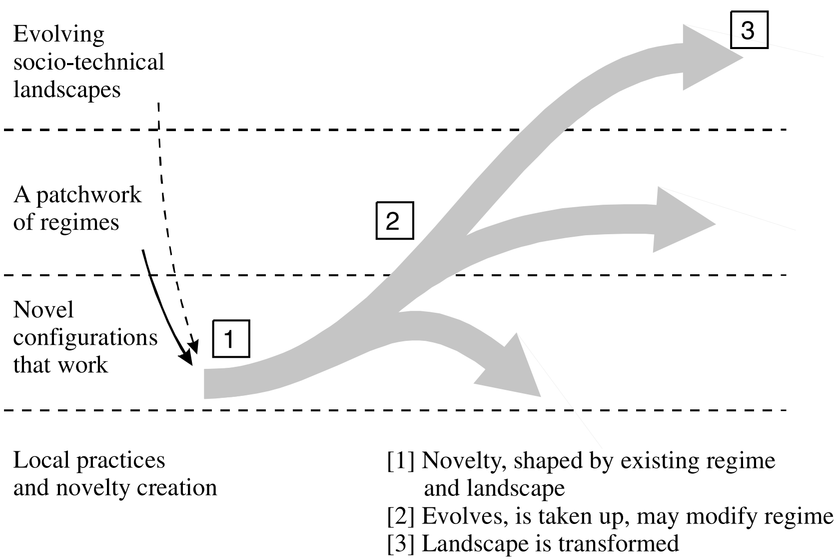


Source: Geels (2002a: 1261).

Figure 3.2 Multiple levels as a nested hierarchy between multiple levels. This leads to several propositions regarding system innovation and technological transition.

First, novelties emerge in technological and/or market niches. Niches are crucial for system innovation, because they provide the seeds for change. The emergence of niches is strongly influenced by existing regimes and landscape, as is represented with downward arrows in Figure 3.3. The downward arrow from the regime level is solid, while the arrow from the landscape level is dotted. This is meant to indicate that the influence from the regimes on niches is stronger and more direct than the influences from landscapes, which is more diffuse and indirect. The novelties are produced on the basis of some existing knowledge and capabilities and geared to the problems of existing regimes. The interpretation of the functionality of novelties occurs on the basis of existing regimes. Novelties start small and insignificant and only later grow out into something large.

Second, the further success of a novelty is 'governed by processes within the niche, by developments at the level of the existing regime and the sociotechnical landscape. So it is the alignment of developments (successful processes within the niche reinforced by changes at regime level and at the level of the sociotechnical landscape) which determine if a regime shift will occur' (Kemp et al., 2001: 277). Diffusion and breakthrough of new



Source: Rip and Groen (2001: 27); courtesy of Edward Elgar.

Figure 3.3 The dynamics of socio-technical change technologies occur as the outcome of linkages between developments at multiple levels. Diffusion and breakthrough thus depend on wider external circumstances at the regime and landscape levels, which create windows of opportunity. This means that novelties may remain 'hidden' in a particular niche, as long as wider regime and landscape processes do not create a window of opportunity.

Third, contingent factors, actor strategies and social mechanisms also play a role. The convergence of processes at multiple levels increases the chance of regime transformation, that is, it creates the 'right' circumstances. Although processes at different levels can converge and create opportunities, the actual linkages always need to be made by actors in their cognitions and activities. Social mechanisms can speed up or accelerate this process.

Although Figure 3.3 has similarities to the standard diffusion curve, it has a different focus. While standard diffusion curves represent (relative) market shares over time, the MLP indicates the changing relationship between technology and activities in local practices, represented as increasing structuration. In the niche phase, actors put in a lot of work to get a new technology on the agenda. Working in precarious networks, they aim to stabilise the technology and create a 'configuration that works'. As a technology gradually moves from the niche to the regime phase, a reversal occurs. Rules stabilise and begin to have strong coordinating and structuring effects on actors in local practices. When technologies become embedded in physical infrastructures and are normal aspects of cultural lifestyles, they become part of wider landscapes. They are very stable, and have a strong force on societal development. Of course, this movement from technological niche to market niche to mass market to regime and even landscape occurs as the technology increases its market share. But although these processes occur in tandem, they have a different focus.

Some Weaknesses in the Multi-level Perspective

Although the multi-level perspective offers relevant insights with regard to system innovation, it also has some weaknesses.

First, the concept of technological regime refers mainly to rules regarding design and production of technology. Van de Poel (1998), for instance, specifically talks about 'design' regimes carried by a technical community, for example, guiding principles, promises and expectations, design criteria, design tools, requirements and specifications. But other groups are also important in socio-technical systems, for example, users, suppliers, policy makers, societal groups and scientists. The concept of technological regime is not wide enough to understand (dynamics in) socio-technical systems.

Second, the perspective (as represented in Figure 3.3) has a bias towards the novelty, sketching its 'innovation journey' through time. The storyline has the sequence of novelty, (re)alignment, reversal, stabilisation, further change. The regime is analysed only with regard to its effects on the novelty, that is, as a barrier to be overcome or as creating windows of opportunity. There is a danger that the analysis follows a David and Goliath storyline, where the niches are the heroes that have to overcome the 'barriers' of the regime. To counter the bias for novelties and heroic storylines, more explicit attention needs to be paid to ongoing processes at the regime level.

Third, although the MLP is developed to understand radical innovations and regime shifts, it is not very explicit about the process of technological replacement. The framework is used mainly to understand the *emergence* of new technologies, but does not say much about its relation with the existing regime, which is not even represented in Figure 3.3.

Fourth, the multi-level framework pays little attention to the diffusion process itself. The strong point is that the framework focuses attention on the wider *circumstances* of diffusion, that is, windows of opportunity, which emerge because of linkages between multiple processes. But a weak point is that the diffusion process from niche level to regime shift is unclear.

Fifth, the thick arrow in Figure 3.3 suggests a focus on *one* promising novelty (for example, electric vehicles, grain elevators). But it seems more likely that system innovations come about through interactions between *multiple* technologies.

Sixth, there has been too little effort, so far, to further differentiate the perspective. More work should be done to refine the MLP, in terms of patterns and mechanisms.

3.3 ADDITIONS TO THE MULTI-LEVEL PERSPECTIVE

This section makes additions to the multi-level perspective to solve the first three problems, mentioned above. The last three problems are addressed in Section 3.4.

From Technological to Socio-technical Regimes

Although the concept of technological regimes is useful for analysing dynamics in engineering communities, it is too narrow to analyse dynamics in socio-technical systems. The elements of socio-technical systems and their linkages are the outcome of activities of a wide range of social groups, for example, engineers and designers, suppliers, scientists, policy makers, consumers and societal groups. Socio-technical systems are stable because the activities of these different groups are aligned to and coordinated with one another. To understand coordination between these social groups, I propose to further widen the scope from technological to socio-technical regimes. Chapter 1 (Section 1.3) described how rules guide not only the activities of engineers and designers but also those of other social groups. When the rules carried by different social groups are linked to one another, this results in coordination of the activities of different social groups. The term 'socio-technical regime' refers to this semi-coherent web of rules. The degree of alignment or tension in the web of rules is an indicator of the stability of linkages in socio-technical systems. In stable systems, the rules are well adjusted, resulting in the alignment of activities. If there are tensions between rules, the activities of different social groups go in different directions, resulting in a weakening of linkages, and possibly windows of opportunity for radical novelties.

The stability of socio-technical systems is of a dynamic kind, meaning that incremental changes in the activities of social groups occur continuously. ST systems remain 'dynamically stable' as long as the activities of social groups and the incremental changes are aligned and go 'in the same direction'. In that case, trajectories or lineages emerge in technical development (and also in policy, user preferences and science). How can we understand this coordination between social groups?

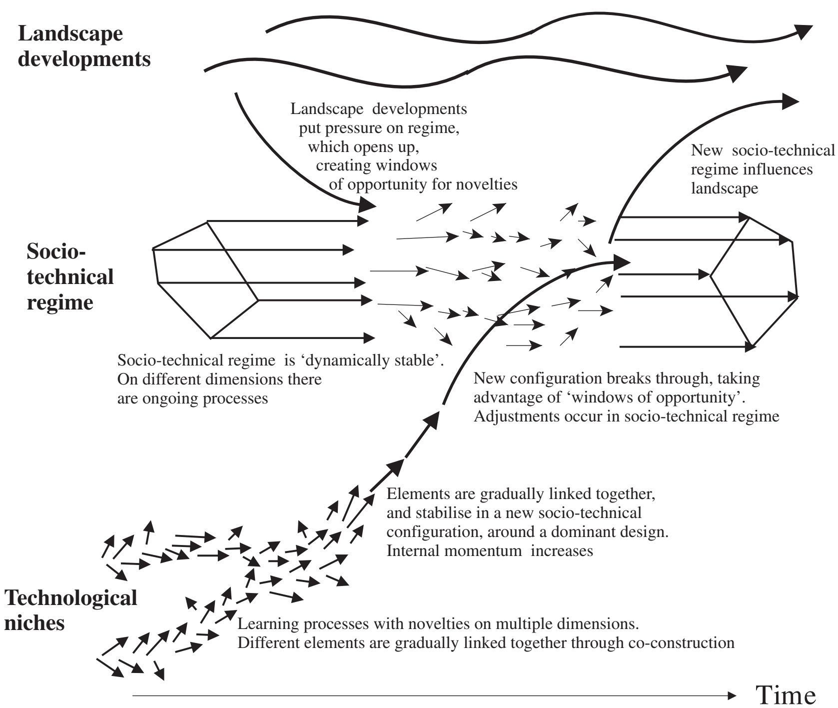
Ongoing Processes at Regime and Landscape Levels

In the literature, the multi-level perspective has been used mainly for understanding the emergence and further diffusion of novelties. To counter the bias for novelties, I argue that more explicit attention needs to be paid to ongoing processes in socio-technical regimes, for example, emergence of new markets, policy dynamics and new technologies which can act as stepping stones. Regimes should be analysed not only as barriers, but also as opportunities.

On the *regime* level there are incremental ongoing processes 'down the design hierarchy' (Clark, 1985). In Chapter 1 (Section 1.3), I distinguished five regimes, which are linked together in socio-technical regimes: technological, science, user/market, policy and socio-cultural. In all these regimes, there are ongoing processes, which result in trajectories. At the end of Chapter 1 (Section 1.3), I represented the ongoing processes in different regimes and their alignment in Figure 1.8. When the steps in the ongoing regime processes go in the same direction, they result in trajectories, which can be represented as straight lines. If we add a slightly adjusted version of Figure 1.8 to the multi-level perspective we get an improved and more dynamic representation, with less bias towards the novelty. In Figure 3.4 I have represented the ongoing processes in different regimes with relatively long arrows. Although the different regimes are linked and co-evolve, they also have internal dynamics. If these internal developments diverge, it results in 'tensions', represented in Figure 3.4 with shorter diverging arrows, indicating uncertainty and differences of opinion. Tensions may lead to periods in which linkages are weakening or 'loosening up'. Such periods form windows of opportunity for radical innovations to break out of their niches. This means that regimes should be analysed not only as barriers, but also as opportunities for novelties.

At the landscape level changes usually take place slowly, for example, cultural changes, demographic trends and broad political changes. The slowly evolving landscape developments are represented in Figure 3.4 with long arrows. Landscape changes may put pressure on the regime.

At the niche level there is much uncertainty and actors may work in different directions, represented by a variety of arrows going in different



Source: Geels (2002: 1263).

Figure 3.4 A dynamic multi-level perspective on system innovation directions. There may even be multiple niches. These niches may compete, but can also gradually link up and reinforce one another, which is indicated in Figure 3.4. Learning processes take place on several dimensions, for example, technology, user preferences, policy measures, infrastructure and symbolic meaning. In successful niches the learning processes are gradually aligned, leading to more stability and convergence. When this leads to a dominant design, developments become more predictable, represented by the arrows growing longer. The innovation may break out of its niche when ongoing processes in the regime and landscape create windows of opportunity. A system innovation occurs when the new innovation conquers wide market shares and links up with ongoing processes in the regime. This is accompanied by wider adjustments in the socio-technical regime. System innovations thus involve not only technology and market shares, but also changes on wider dimensions such as regulation, infrastructure, symbolic meaning and industrial networks. Over time, the new regime may influence wider landscape developments.

The basic notion of the improved multi-level perspective is that system innovations are the outcome of linkages between developments at multiple levels. An important consequence of my addition is that the existing regime should not (only) be seen as a barrier for radical novelties, but the ongoing processes in regimes may also create opportunities. If radical novelties can link up with 'ongoing processes' in the regime, they can break out of the niche level. The 'linking up' of a radical innovation may occur in any of the dimensions of the socio-technical regime, for example, in the established technology as an auxiliary device (as add-on), new regulations, cultural debates, newly emerging markets, scientific insights, strategic games in industrial networks and so on.

An important aspect of the MLP is to do away with simple causality in system innovation. There is no single 'cause' or driver of system innovation. Instead, there are simultaneous processes at multiple dimensions and levels. System innovations come about when these processes link up and reinforce one another ('circular causality'). Such innovations are not simply 'caused' by novelties at the niche level, because their breakthrough and diffusion depends on processes and circumstances at the regime and landscape levels. In fact, most radical innovations fail to break out of the niche level. When the socio-technical regime is stable, radical novelties usually have little chance. But even when there are tensions in the regime, there is no guarantee that novelties will break through. In those circumstances, product champions and niche advocates may have high expectations and promise a glorious future for their novelties. But it is very possible, and even likely, that the problems will be solved with incremental innovations within the regime and that niches will fail (Figure 3.5).

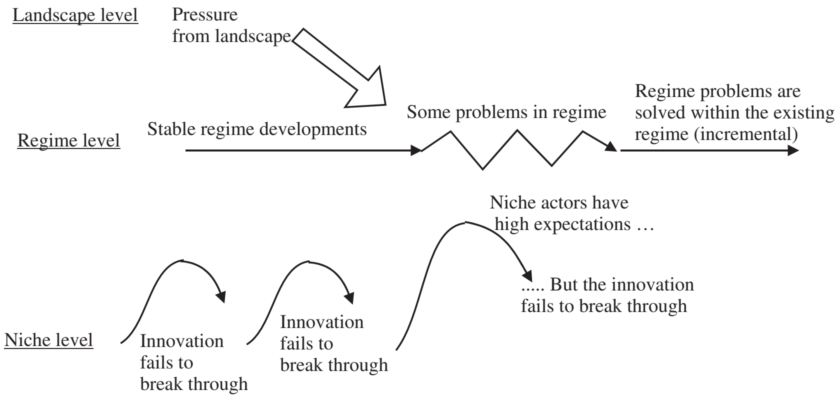


Figure 3.5 Failures of niche innovations

Phases in System Innovation

I propose four phases in system innovations (these phases are somewhat different from the four phases distinguished in Rotmans et al., 2000, 2001). In the first two, sociological and socio-technical processes are more important, and the emphasis is on perceptions and rules and the alignment of heterogeneous aspects (co-construction). In the last two, economic competition plays a more important role. Economic competition and replacement thus appear as a particular phase in system innovations.

The first phase encompasses the emergence of novelty. A new technology is not born in an empty world, but in the context of existing regimes and landscapes. Novelties often emerge to solve small problems in the existing regime. It may even be unclear what functions the novelty will fulfil. In that sense, a novelty is a solution looking for problems.

A novelty emerges in technological niches and small market niches. It is carried by a precarious network of actors on the basis of expectations and 'diffuse scenarios'. Because there is much uncertainty about the best technical design, actors improvise on the basis of design rules from the existing regime and engage in technical experiments to work out the best design and find out what users want. There may be aspects of continuity with regard to design rules and ideas about functionality. Under the sub heading of socio-cognitive approaches (Chapter 2, Section 2.3) I noted that design and functionality of new technologies are often initially interpreted with rules and categories associated with the old technology. I cited several historians who gave historical examples of this dynamic. The early ideas about design and functionality may be very different from the eventual outcomes.

This means that the notion of ‘radical’ innovation may well be *ex post*. In their early phase the revolutionary potential of novelties may not even be recognised.

The second phase takes place in small market niches. Technical specialisation and exploration of new functionalities may occur. The social network that supports the novelty may develop into a dedicated community of engineers and producers. Their activities aim to improve the new technology and find out more about user preferences, legislation and so on. These activities have the character of probing and learning, working outward from established practices to explore new ways: ‘New practices do not so much flow directly from technologies that inspire them, as they are improvised out of old practices that no longer work in new settings’ (Marvin, 1988: 5). The new technology is gradually emancipated from the old regime, because new technical groups themselves specialise. Professional associations and special journals are created, as well as new places for communication. Engineers and designers meet at conferences, establish problem agendas, and discuss promising findings and search heuristics. The learning processes in technical communities gradually result in new design rules. Guiding principles, performance specifications and search heuristics shift to the background, forming a stable frame for economic (inter)actions and calculations. This is a sociological explanation of lock-in and emerging path dependence. As the rules stabilise, the new technology develops a technical trajectory of its own. The performance of the new technology gradually improves. This technical emancipation and specialisation only occurs if there are sufficient resources. While the first phase is carried by technical pioneers or dedicated R&D projects, the second phase requires commercialisation in market niches or government subsidies.

Through concrete interactions with the technology, users gradually learn about the technology, how it can be used and what they want from it. Users initially make sense of new technologies with perceptions and concepts related to established technologies and practices. Gradually they develop new concepts and interpretations (Clark, 1985: 234–45). An important aspect of the second phase is the stabilisation of rules, which is a pre-condition for the wide diffusion of new technologies (Spar, 2001).

The third phase is about breakthrough of the new technology, wide diffusion and competition with the established regime. While the first two phases occurred in particular niches and were relatively ‘invisible’ to regime actors, the wider diffusion gives the new technology high visibility. With the breakthrough in mainstream markets, the new technology enters a competitive relationship with the established regime. There are two complementary explanations to understand the dynamics in this phase: (a) external circumstances and (b) internal ‘drivers’.

The first explanation concerns external circumstances. The multi-level perspective highlights the point that wide diffusion of novelties depends on circumstances and windows of opportunity. This point was also emphasised by some authors in the literature review, for example, Freeman and Perez (1988), Staudenmaier (1989) and Summerton (1994). The following circumstances are important to create windows of opportunity for the wide diffusion of novelties:

1. Internal technical problems in the existing regime that apparently cannot be met by the available technology. Rosenberg (1976) coined the term 'bottlenecks', while Hughes (1987) talks of persistent 'reverse salients'. Another driver to look for alternative technologies is 'diminishing returns of existing technology' (Freeman and Perez, 1988). Not only existing problems, but also *expected* problems are important. Constant (1980) used the term 'presumptive anomalies' for those problems that are expected on the basis of scientific research. It is not just the existence of technical problems, but also the shared perception and placement on problem agendas which is important.
2. Problems external to the system. Examples are negative externalities such as environmental impacts, risks and concerns about safety. The externalities may be due to unforeseen complexities (for example, risks), or be caused by strong growth and expansion of the technology. Regime actors tend to downplay negative externalities, because they do not hamper the internal functioning of existing regimes. The externalities are often picked up and placed on the problem agenda by 'outsiders' such as societal pressure groups (for example, Greenpeace), outside engineering and scientific professionals, or outside firms (Van de Poel, 2000). To get negative externalities on the technical agenda of firms and designers, there may be a need for consumer and political pressure, and regulatory measures.
3. Changing user preferences create opportunities for the emergence of new technologies when established technologies have difficulties meeting them. Changing user preferences may lead to the emergence of new market niches. If new technologies link up with these niches, they may ride along with their growth (piggy back).
4. Strategic and competitive games between firms may create opportunities for new technologies. Investments in a new technology may be a move in strategic games, trying to leap ahead of other firms that stick to the old technology.
5. Availability of complementary technologies for the new technology. A technology may break out of its niche when complementary technologies make wider application possible.

These circumstances in the regime may be influenced by wider landscape developments. For instance, wide cultural changes may lead to new values, which influence user preferences. Or certain problems may receive much societal attention (for example, hygiene at the end of the nineteenth century, or the environment since the 1970s), which places negative externalities higher on the problem agenda. Or pervasive technologies may provide opportunities for novelties to link up with.

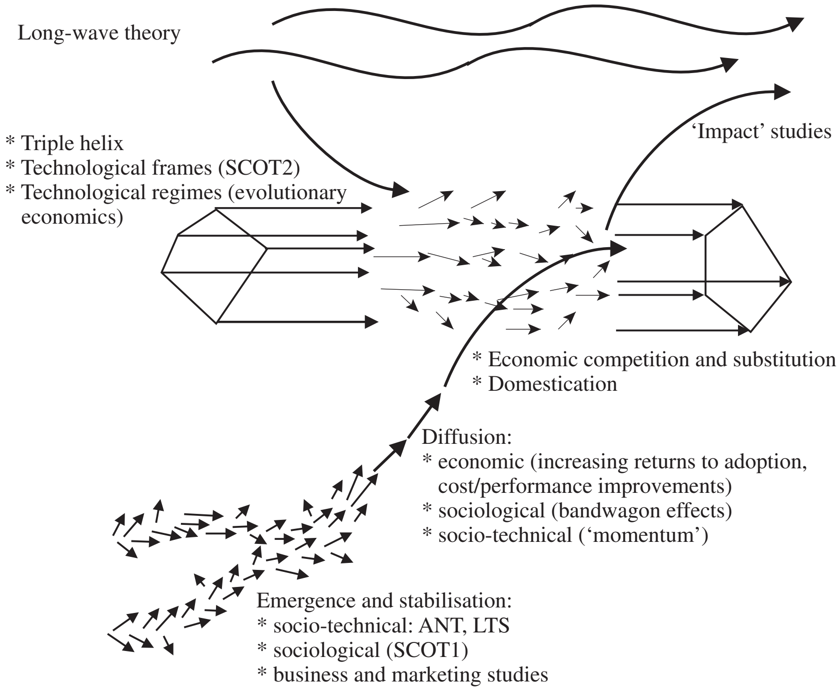
Besides such external circumstances at the regime level, there are also internal 'drivers' which stimulate the diffusion of novelties. Disciplinary perspectives highlight different drivers, focusing on different important aspects. *Economic perspectives* argue that the improvement of the price/performance ratio of the novelty is a crucial internal dynamic, which drives diffusion. Economic calculations are possible because a stable frame has been formed in the earlier phases. Improvements in price/performance may result from improvements in the product or in the production process (learning by doing). It is also possible that 'increasing returns to adoption' kick in, as argued by economic path-dependence theorists. That could be due to learning by using, network externalities, scale economies in production, informational increasing returns and technological interrelatedness. In *socio-technical perspectives* diffusion is understood as a process of creating linkages between heterogeneous elements:

> What is usually called technology diffusion is really a transformation process by which a new technological regime grows out of the old regime. Technology adoption is an active process, with has elements of innovation in itself. It is connected with the availability of new technologies, with expectations, new skills, management systems, new supplier–user relationships, changes in the regulatory framework and new ideas. Behaviors, organization and society have to re-arrange themselves to adopt and adapt to, the novelty. Both the technology and social context change in a process that can be seen as co-evolution. (Rip and Kemp, 1998: 389)

The diffusion of technologies is possible *because* more elements are linked together around an emerging innovation. The internal drive for diffusion is explained by the notion of 'momentum' (Staudenmaier, 1989; Hughes, 1994). Specialised groups emerge which act to protect and expand their economic interests. New infrastructures may be created, designed and maintained by new governmental departments and agencies. Engineers set up professional organisations and dedicated engineering schools. But momentum refers not only to the mass of social and institutional groups, but also to the 'hardness' of artefacts, special-purpose machines, factories and physical infrastructures. The increasing number of linkages leads to irreversibility, mutual dependencies and path dependence. While technological development is strongly shaped by society in the early phases, it becomes a force of its own in later phases, shaping further societal developments. In the *sociological* literature we find contributions which note the importance of mechanisms in the diffusion process. This is an uneven process that proceeds in fits and starts. There are periods of acceleration and slowing down as a result of the actions and interactions of social groups. Perceptions, expectations and strategies can change suddenly, as a result of social mechanisms. Some mechanisms have been identified in the literature review, for example, hypes, strategic games, bandwagon effect and innovation race, social struggles, importance of outsiders and the sailing ship effect (see Chapter 2, Section 2.3). Although some mechanisms have been identified, no systematic work has been done on the issue. I aim to take a first step in that direction, using my case studies to construct a list of mechanisms (see Chapter 7).

The fourth phase is that of gradual replacement of the established regime and wider impacts on society. For several reasons, the new technology gradually replaces the old. First, the cost/performance ratio of the new technology is improved gradually, as a result of incremental innovations. Second, societal domains can consist of many market niches, with different selection criteria. The new technology then needs time to conquer these market niches. Third, the creation of a new socio-technical regime may require a wide range of transformations, for example, new infrastructures, new user practices, new policies and new organisations. Such adjustments can take time. Fourth, the capital intensity of existing technologies may have a delaying effect. If firms have sunk investments in complex production systems, they tend to stick to the old regime until investments are written off. Because of all these reasons, old technologies may hold on to particular market niches for a long time. Old and new technologies may thus co-exist for a substantial period, before the old technology is entirely replaced. In this phase the new system may have wider impacts on other societal domains. Impacts on the macro level cannot be attributed to single causes or a set of factors, but emerge in the process itself. Impacts are co-produced, when the new system links up with ongoing developments in other domains or at the landscape level.

In the course of these four patterns, the relationship between technology and society gradually changes. In the early phases, new technologies are strongly shaped by their environment. But as more elements are linked together, the new technology acquires more momentum. Then the new technology gradually becomes a shaping ‘force’ on its environment. These four phases also allow a pragmatic positioning and integration of insights from different literatures. This is schematically represented in Figure 3.6.



Source: Geels (2004b).

Figure 3.6 Integration of different literatures in the multi-level perspective

3.4 PATTERNS IN SYSTEM INNOVATION

This section addresses the three remaining problems mentioned at the end of Section 3.2. One problem was that the multi-level perspective does not pay much attention to the diffusion process. This problem has been partly addressed above, when describing the third phase in system innovation. But more can be said about it. Another problem was the focus on *one* radical innovation, instead of *multiple* technologies. A third problem was that there has been little effort, so far, to further differentiate the multi-level perspective. To deal with the last problem, this section proposes several patterns in system innovation, that is, characteristic dynamics that stretch over different phases of the entire system innovation process. These patterns aim to solve the other two problems. Three patterns will be described: (a) fit–stretch pattern in the co-evolution of form and function, (b) co-evolution of technologies and (c) diffusion as a trajectory of niche accumulation.

Fit–Stretch Pattern in the Co-evolution of Form and Function

In Chapter 1 (Section 1.3), it was shown that technological artefacts have a dual nature. On the one hand they have a technical form (represented by design rules, search heuristics and technical characteristics); on the other they fulfil a social function (represented by user preferences and performance requirements). This is close to the dimensions used by Abernathy and Clark (1985) for their innovation typology: (i) technology choice and design, and (ii) user environment. With regard to the form and function, we can look more closely at the relationship between niche and regime during system innovations. On the form and function dimensions, a niche can either fit with the existing regime, or stretch from it: fit means that the niche rules are very similar to the regime rules; stretch means that the niche rules deviate from the regime rules. If we cross these two dimensions, we get four quadrants with different experimentation strategies with new technologies in niches (Table 3.1):

- Selective substitution means that both the chosen technology and the targeted use environment remain close to the existing regime. For experiments with electric vehicles, this would mean that existing all-purpose automobiles or delivery vans are retrofitted with an electric motor and battery.
- Market differentiation means that the technical form remains close to the regime, but that the targeted function departs from the regime. For instance, electric bicycles consist of combinations of existing technologies, but explore new markets.
- Leapfrog design for substitution means that the targeted function remains close to the regime, while the technical form differs strongly. Fuel cell electric vehicles are an example, because the technology is very different, but the market is the all-purpose sedan.

Table 3.1 Typology of emergent market introduction strategies

Technical form / Use environment	Fit	Stretch
Fit	Selective substitution	Leapfrog design for substitution
Stretch	Market differentiation	Exploration of new regime

Source: Hoogma (2000: 21).

- Exploration of a possible new regime means that both technology and targeted use environment differ considerably from the existing regime. Examples are lightweight electric vehicles, consisting of plastic bodies with different designs, used for short-distance journeys (for example, 50 km to work). Electric station cars also entail a different mobility practice. An electric station car is an electric vehicle used by commuters to drive small distances to railway stations, from where they continue their journey by rail (chain mobility). In this function, the relatively short range of electric vehicles is not a problem.

Using these distinctions we can propose a particular pattern in the co-evolution of form and function in system innovations. The idea is that new technologies start their life as a fit–fit strategy. Both the form and the functionality of the niche are close to the existing regime. Gradually, new design possibilities and new functionalities are explored, resulting in stretch–stretch strategies. The new technology develops a form of its own, with its own set of design rules. New markets and functionalities may also be explored. I have sketched this pattern in Figure 3.7, and also indicated the four different phases in system innovation.

Hoogma (2000) used this idea to investigate niche experiments with electric vehicles in the 1980s and 1990s, in several countries: Switzerland, Germany, France, the United States, Norway and Japan. His findings suggest that niche experiments moved from fit–fit to stretch–stretch. Most experiments in the late 1980s took place as selective substitution, remaining

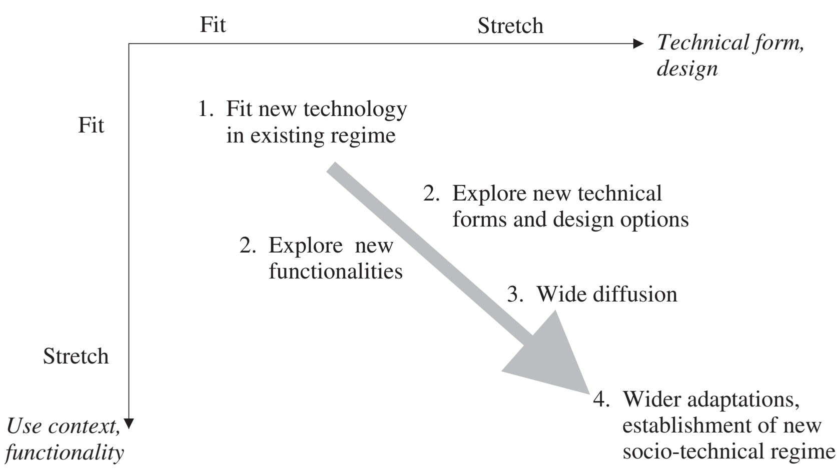


Figure 3.7 Fit–stretch in co-evolution of form and function

close to the form and function of the automobile regime. Examples are: (a) existing vans, retrofitted with electric batteries and an electric motor; (b) special utility vehicles (for example, mail delivery); (c) fleet vehicles; (d) electric vehicles in holiday resorts; (e) electric buses/trucks; (f) golf carts; and (g) self-conversions, pioneers who convert existing cars to electric vehicles themselves. Some experiments moved in a somewhat stretch direction, for example, new markets for electric bikes were explored in Japan, and the CitySTROMers, a converted Golf model for multiple ownership, was investigated in Germany. But in the late 1990s there was much more variety in experimentation strategies. Several experiments explored new directions in form and function, for example, small lightweight electric vehicles in Switzerland and Germany; self-service electric vehicles in Japan, France and Switzerland; and electric station cars in the United States. There were also experiments that mainly explored new technical directions, for example, experiments with starter generator hybrids and advanced battery cars in Germany and gas turbine hybrids in France; and experiments that explored new functions and markets, for example, electric bikes in Japan and Switzerland, and electric scooters in the United States.

Hoogma thus concluded that experiments with electric vehicles gradually moved from fit–fit to stretch–stretch. But his findings cover only a brief period (15 years). And the electric battery vehicle niche collapsed in the beginning of the twenty-first century, as fuel-cell vehicles became more popular, drawing subsidies and resources away from electric battery vehicles. Therefore it will be interesting to see whether the fit–stretch pattern holds for a longer period during system innovations.

Co-evolution of Multiple Technologies

In innovation studies and science and technology studies, the focus has usually been on the emergence and diffusion of *one* technology, for example, electric vehicles, bicycles or grain elevators. But to understand system innovation, linkages and interactions between *multiple* technologies are important. There is co-evolution between multiple technologies. From different literatures I derived several kinds of interaction.

One is the interlocking, alignment and positive feedback among technologies, especially in the diffusion phase. Rosenberg (1982), for instance, argued for the importance of interrelatedness between technologies and the need for a systems perspective:

> The growing productivity of industrial economies is the complex outcome of large numbers of interlocking, mutually reinforcing technologies, the individual components of which are of very limited economic consequences by themselves.

> The smallest unit of observations, therefore, is seldom a single innovation but, more typically, an interrelated clustering of innovations. . . . The importance of these complementarities suggests that it may be fruitful to think of each of these major clustering of innovations from a systems perspective. (p. 59)

In my discussion of long-wave theory I mentioned Ayres (1989), who drew attention to the importance of cross-sectoral clustering between technologies. For instance, it was the combination of steel, gasoline and the internal combustion engine which made large-scale production of the automobile possible. Grübler and Nakićenović (1991) also argue that the diffusion process needs to be understood as a clustering of innovations:

> The diffusion processes within each cluster constitute families of interrelated systems that enhance each other. In other words, each diffusion cluster is a network of interrelated families of socio-technical systems that reinforce and build upon each other. . . . One successful diffusion has a positive feedback and catalytic effect on the development of many others within the whole cluster. (p. 337)

In their discussion of the literature on diffusion, Lissoni and Metcalfe (1994) conclude that diffusion should be understood as a multi-technology process, involving linkages and co-development:

> Compatibility, inter-relatedness and co-development are emerging as important themes in modern diffusion research. Furthermore, the single innovation is no more seen as the appropriate unit for diffusion analysis. Rather what is being diffused is often a sequence of innovations with an evolving design configuration which itself develops in response to competing and complementary configurations. A multi-technology approach is called for. (p. 107)

So there is good reason to expect that system innovations do not come about through the breakthrough of one technology, but that they involve combinations of multiple technologies.

A second interaction is complementarities between technologies, which forms a particular form of interlocking. Complementarities means that a given invention, however promising, cannot fulfil its full potential unless other inventions are made which relax or bypass constraints, which otherwise hamper its diffusion and expansion. Rosenberg (1976) gives the example that steel was important for the wide diffusion of trains:

> The critical importance of steel rails to growth in railroad productivity was that they were far more durable, lasting more than ten times as long as iron rails, and that they were capable of bearing far heavier loads than iron rails without breaking. (p. 202)

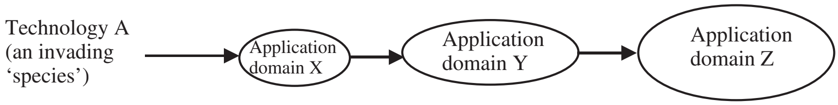
A third interaction is the cumulative effect of many incremental innovations to substantially improve a product's performance. From the complex products and systems literature (see Chapter 2, Section 2.2), we know that technological products are made up of a range of subsystems and components. Their structure is a technical hierarchy. Before a new technology diffuses widely, a range of incremental innovations and improvements in subsystems and components is needed to improve its performance. Rosenberg mentions some of the incremental innovations which improved the functioning of trains:

> The cumulation of small innovations and relatively modest design changes brought about, between 1870 and 1910 alone, more than a tripling of freight car capacities with only a small increase in dead weight . . . The greater loads and speeds made possible by the rolling stock could not have been achieved, however, without several other significant inventions: the control of train movements through the use of the telegraph, block signalling, air brakes, automatic couplers, and the substitution of steel rails for iron. (ibid.: 201)

I conclude that there are some incidental examples and insights in the literature with regard to the co-evolution of technologies. But I also conclude that little systematic research has been done on this topic. In this book I put the issue on the agenda, and I shall return to the subject in Chapter 7, using my case studies to examine it further.

Diffusion as a Trajectory of Niche Accumulation

Diffusion is a complex process, involving changes on many dimensions. I have already described how the multi-level perspective and economic, sociological and socio-technical approaches highlight different aspects of the process. But one aspect of the multi-level perspective has still remained unclear: how do radical innovations move from the niche to the regime level? To answer this question and further contribute to an understanding of diffusion, I can use a building block from the literature review on evolutionary economics (Chapter 2, Section 2.3). Both Levinthal (1998) and Schot (1998) suggested that technological diffusion does not occur all at once, but in subsequent steps. The critical factor for diffusion is whether the new technology remains in the initial small niche, or is able to penetrate a broader set of niches. Following Levinthal (1998) and Schot (1998) I propose that the general pattern by which radical innovations move from the niche to the regime level is that they follow *trajectories of niche accumulation*. The breakthrough from niche to regime level occurs gradually, as a new technology 'branches' or 'penetrates' different application domains (see Figure 3.8).



Source: Levinthal (1998: 243); courtesy of Oxford University Press.

Figure 3.8 Diffusion as accumulation of niches

A simplification of Figure 3.8 is that it suggests that the niches are waiting 'out there' to be entered by new technologies. This is not always true, since many niches depend on the dedicated construction work of product champions and on windows of opportunities created by external circumstances. Another simplification is that Figure 3.8 has a focus on technologies and markets, and does not show the other dimensions of socio-technical systems, for example, policy, symbolic meaning and infrastructure.

3.5 CASE-STUDY PROTOCOL

The next three chapters present the following three historical case studies about system innovation, focusing on the leading countries:

1. The transition from sailing ships to steamships in Britain.
2. The transition from horse-drawn carriages to automobiles in America.
3. The transition from piston engine aircraft to jetliners in America.

Although the case studies are named on the basis of the main technological changes, they are all about major shifts in socio-technical systems, involving changes in user practices, infrastructure, legislation, supply networks, science and cultural values.

The aim of the case studies is to test the conceptual perspective. The improved multi-level perspective gives a basic answer to the first research question. Systems innovations occur because of linkages between processes at multiple levels, in particular niche, regime and landscape levels. Radical novelties emerge in technological niches, which are shaped by the wider context of the socio-technical regime and landscape. Diffusion and breakthrough of novelties depend on their linking up with ongoing processes at the regime and landscape levels, which creates windows of opportunity. The second research question was answered by the three patterns described above. The third research question about mechanisms has not yet been answered, although some mechanisms have been identified in the literature review (see Chapter 2, Section 2.3). I shall use the case studies to address this question in Chapter 7.

To test the theoretical answers to the research questions, each case study has the same structure. In an introductory section I give a first quantitative mapping of the entire process, often in terms of market shares of old and new technologies over time. This mapping suggests that the dynamic is one of technological substitution, which is too simple. But as a first step it is useful, because it gives an idea of the time period of the transition, and possibly some interesting turning points (slow-downs, accelerations). The introductory section also describes the elements of the relevant socio-technical system in the form of a schematic picture. It was not possible to sketch both the old and new socio-technical systems in one picture. I have chosen to represent the outcome of the transition process, that is, the new system, because that gives an indication of the elements that will be woven together in the course of the empirical description. I also present a figure of the main social groups involved in maintaining (and changing) the elements of the socio-technical system. As indicated in Chapter 1 (Section 1.3) the socio-technical system, social groups and the socio-technical regime are always interrelated in practice. The sketch of the socio-technical system and social groups is not exhaustive. The main purpose is to give the reader a first idea of the elements and actors that play a role in the story.

The next four sections describe the system innovations, using the four phases as a structuring device. Each section describes dynamics at the three levels of the multi-level perspective, working from landscape to regime to niche. For each section, I shall describe the following dynamics:

- What were the relevant external landscape developments?
- What were the main developments in the socio-technical regime? I shall pay attention to technological developments, market dynamics, policy actions, problems and tensions, and also show how these regime processes were influenced by landscape developments.
- Which novelties emerged in which niches? Which actors were involved? What problems did they encounter and how did they try to overcome them? How were niche dynamics related to ongoing processes at the regime and landscape levels?

Each case study ends with a concluding section, where the following questions are explicitly addressed. How did niches emerge in the context of the existing regime? To which ongoing processes at both regime and landscape levels did the niche link up? These two questions are proxies to test the usefulness of the multi-level perspective. The following questions test the proposed patterns. How did multiple technologies interact? How did diffusion occur as a trajectory of niche accumulation? Did the four phases indeed occur in the system innovation? Was there a fit–stretch pattern in the co-evolution of form and function?

NOTES

1. An analogy can be made with neural networks in the human brain. Connections in neural networks grow stronger or weaker as they are used more or less frequently. Connections may even die out if not used at all. Likewise, rules change as they are used. Rules that are no longer used gradually die out.
2. For example, when the telephone was invented in 1876, the person on one end of the line had to shout loudly while the person on the other end had to listen carefully. Cars with internal combustion engines in the 1890s were dangerous and dirty machines that frequently broke down. The ENIAC computer, completed in 1946, weighed 30,000 kg and consisted of 48,000 thousand switches. These switches frequently broke down and had to be replaced (Nijholt and van den Ende, 1994).

4. The transition from sailing ships to steamships in British oceanic transport (1780–1890)

4.1 INTRODUCTION

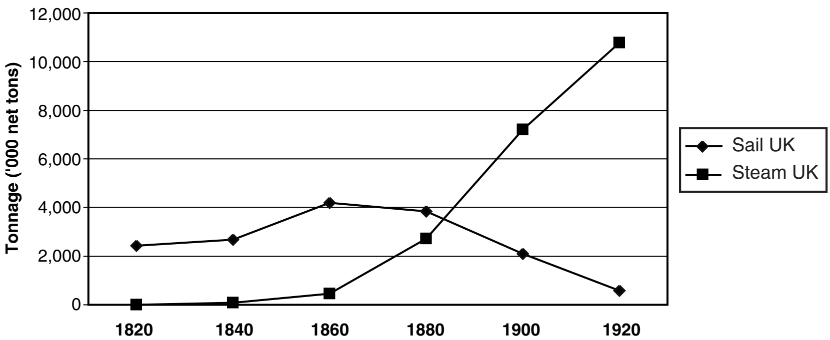
In the nineteenth century, Great Britain was the dominant shipping nation, with a fleet larger than that of other European countries (see Table 4.1).

Table 4.1 European merchant shipping fleets (1000 net British tons)

	Great Britain	Germany	Netherlands	France	Norway
1780	882	123	398	729	386
1820					
Sail	2,436	–	–	–	–
Steam	3	–	–	–	–
1840					
Sail	2,680	–	–	653	–
Steam	88	–	–	10	–
1860					
Sail	4,204	754	485	928	–
Steam	454	23	11	68	–
1880					
Sail	3,851	927	264	642	1,461
Steam	2,724	177	64	278	58
1900					
Sail	2,096	584	78	510	1,003
Steam	7,208	1,319	268	528	505
1920					
Sail	584	288	24	433	204
Steam	10,777	1,546	969	1,085	1,199

Note: *For some countries before 1860, data are missing.*

Source: Ville (1990: 68–71).



Source: Based on data from Ville (1990).

Figure 4.1 Tonnage of steamships and sailing ships in Great Britain

Great Britain was also the leading country in the transition from sailing ships to steamships. This technological transition is represented in Figure 4.1. It is sometimes portrayed as a simple technological substitution process (for example, Grübler, 1991, discussed in Chapter 2, Section 2.3), but this chapter will show that the transition was more complex. Of course, the transition involved technical changes and innovations in shipbuilding and design practices. But it also involved new ship management and operating practices (for example, centralised fleet management), new functionalities (line services) and new infrastructures (deeper and bigger ports, coal bunkers). Hence, the technological transition was part of a wider system innovation. This chapter analyses how the transition and system innovation came about.

Figure 4.1 gives some indications of the transition process. First, it is clear that steamships existed for a long time before they were widely diffused and replaced sailing ships. In 1850, the percentage of British steamships in terms of registered tonnage was 5 per cent. By 1860 this had grown to 10 per cent, and 51 per cent was reached in 1883. Second, sailing ships were not immediately replaced by steamships, but continued to be used well into the twentieth century. Third, the emergence of steamships occurred in a growing shipping market. This means that the initial growth of steamships was not at the expense of sailing ships.

To analyse the transition to steamships and the wider system innovation, I shall use the conceptual perspective described in Chapters 1 and 3. First, I shall give an indication of the socio-technical system in shipping, to get an impression of the relevant elements. Figure 4.2 presents a sketch of this system, filled in for steamships. Many of these elements did not exist, but

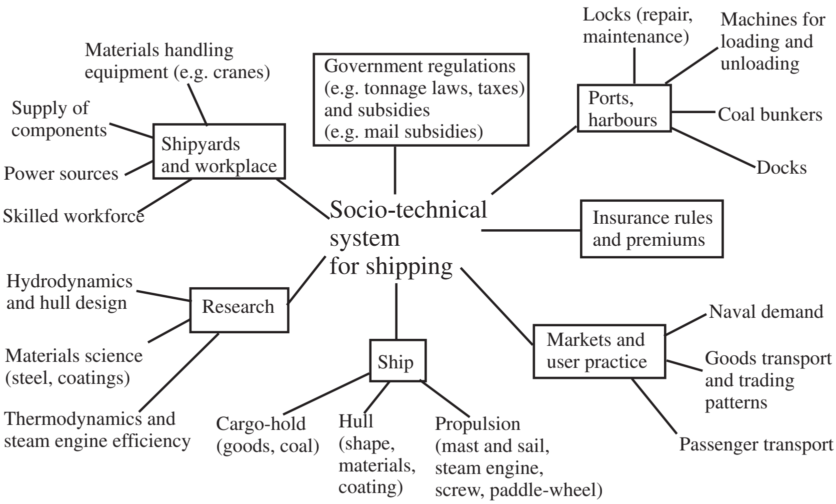


Figure 4.2 Elements and linkages in the socio-technical system for oceanic shipping

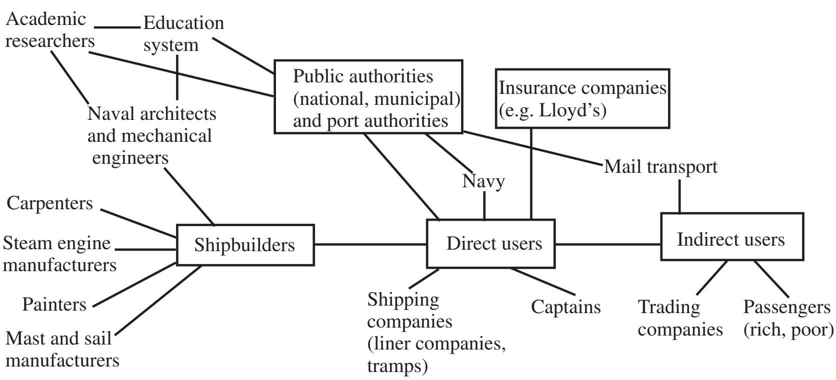


Figure 4.3 Social groups involved in oceanic shipping

had to be created and linked together. I shall analyse how the elements and the linkages interacted and co-evolved.

I also present a sketch of the main social groups involved in oceanic shipping (Figure 4.3). These social groups maintained (and changed) the elements of the socio-technical system. They acted in the context of a set of rules, the socio-technical regime, which involved design rules, problem agendas, search heuristics, exemplars, tax rules, insurance rules, user practices and so on.

These indications of system, social groups and regime show that the transition was complex, involving many elements, actors and rules. The question is how interactions between them resulted in a system innovation. To structure the analysis I shall use the multi-level perspective, described in Chapter 3. This means that the analysis not only pays attention to the regime and system levels, but also to the niche level, where novelties emerge. We shall see that steamships emerged in particular niches, namely on canals and in ports. In ports, early steamships were operated as towboats to help manoeuvre large sailing ships, with which they thus had a symbiotic relationship. The third level is the wider, external landscape. At this level, the nineteenth century was characterised by several macro developments. One such development was industrialisation, which led to an increased demand for raw materials, creating growth markets for shipping. This was facilitated by the rise of political liberalism and a liberalisation of trade (reduction of import tariffs). Another development was large-scale emigration from Europe to the United States, creating another growth market for ships. The hypothesis is that the interaction among these three levels explains the timing and course of the system innovation. This hypothesis will be tested in the concluding section of the chapter.

Another aim of the chapter is to test the patterns described in Chapter 3: co-evolution between multiple technologies, diffusion as a trajectory of niche accumulation, phases, and fit–stretch pattern in the co-evolution of form and function. These patterns will be explicitly addressed in the concluding section.

The empirical description is organised along the four phases in system innovations. Together with an additional section on the pre-history, the story is told in five sections, addressing different periods. The first period (1780–1830) is characterised by technical and institutional innovations in the sailing ship regime, and the emergence of steamships in a particular niche, inland waterways (Section 4.2). The second period (1830–48) is characterised by the entry of steamships in oceanic shipping (Section 4.3). The early steamships were used to improve the functioning of the sailing ship regime, for example, as steam tugs and as mail steamers. In the third period (1848–69) steamships took off in the niche of passenger transportation, linking up with the wider development of European emigration (Section 4.4). The opening of the Suez Canal in 1869 signals the start of the fourth period (1869–90), because it represents the diffusion of steamships in freight shipping (Section 4.5). In this period a new socio-technical regime was created with a worldwide coal infrastructure, bigger and more efficient ports, new shipbuilding practices and new management practices in the operation of steamships. In the fifth period (1890–1914) steamships began to have wider impacts, for example, globalisation of trade, enhanced European emigration, eating habits and agricultural practices (Section 4.6). In this period Great Britain and Germany become entangled in an innovation race, competing for the fastest and largest ships on the Atlantic Ocean. This innovation race sparked off the birth of giant steamships. The chapter ends with an analysis and conclusions (Section 4.7). For each section the following dynamics are described:

- Relevant external landscape developments.
- Main developments in the socio-technical regime, for example, technological developments, market dynamics, policy actions, problems and tensions. How were these regime processes influenced by landscape developments?
- Emergence and development of novelties in niches, and the actors involved, the problems they encountered, and their attempts to overcome them. How were niche dynamics related to ongoing processes at the regime and landscape levels?

4.2 PRE-STEAM DEVELOPMENTS IN OCEANIC SHIPPING (1780–1830)

Landscape Developments

An important landscape development in the late eighteenth and early nineteenth centuries was the construction of an interconnected network of inland waterways. To stimulate trade, rivers were deepened and artificial canals were constructed. This was the period of the 'canal boom'. In Britain the first canal boom was in the 1760s, followed by a second one in the 1790s (Ville, 1990). The first American canals were built in the 1780s, and a veritable canal craze set in after the Erie Canal was completed in 1825. In France, canal construction was concentrated around the 1815–50 period. The improvement of the inland waterways was an important step towards the creation of national markets, as goods could be transported more easily and cheaply over longer distances. Canals offered a transport option 50 to 75 per cent cheaper than roads (O'Rourke and Williamson, 1999). The canal boom provided a good context for the emergence of steamboats (see below).

In America, canals and inland waterways played an important role in another landscape development, namely westward settlements. Pioneers travelled westward, creating settlements and farms as they went along. The inland waterways were important for delivering industrial goods and products, and taking produce to eastern cities. Steamships diffused in America because they linked up with these westward settlements.

Wars were another macro development. Before the American War of Independence (1776–83), American colonies provided cheap timber to Britain. After the war, Britain punished America by denying access to British and colonial markets. Hence, American ship owners and traders had to find alternative shipping markets, for which they used a new and fast sailing ship, the Baltimore clipper (see below). The French Wars (1789–1815) also influenced shipping, pushing up freight prices. These higher prices attracted entrepreneurial ship owners, resulting in a doubling of the registered European tonnage (Ville, 1990). When the war ended, there was an oversupply of ships, causing the European shipping depression of the 1820s and 1830s. This context was unfavourable to sailing ship innovations.

Dynamics in the Sailing Ship Regime

With regard to sailing ships, American shipbuilders and ship owners were more innovative than their European counterparts, who suffered from oversupply and protectionist rules in colonial trade. Many European countries created trade monopolies, restricting colonial trade to their own ships (for example, the British Navigation Acts). Banned from British markets, American traders and ship owners turned their attention to the Atlantic, Mediterranean islands, Mauritius and China. In the small-volume, high-value Chinese trade (for example, in opium and silk), American merchants became competitors of the British East India Company. In these market niches they needed fast ships to evade patrols. The innovative 'Baltimore clipper', a fast but relatively small ship, was the answer. Its hull was not flat-bottomed and full-bilged, but had considerable deadrise, that is, the hull angling sharply up from the keel (Calhoun, 1973). It was also light, making it faster than comparable British models. Baltimore clippers were popular in application domains which needed speed and manoeuvrability more than cargo capacity, for example, smuggling, piracy and the slave trade.

British shipbuilding was guided by different rules and developed in different directions. Because of protectionism, British ship owners enjoyed monopolies in the colonial trade, for example, the East India Company. As a result, British ship owners were more interested in ships with a large cargo-holding capacity than in ships with high speed. This resulted in wide, heavy and sluggish ships. The design heuristics were also influenced by government regulations, in particular the Tonnage Laws of 1773, which specified methods to determine the tonnage of a ship, which in turn determined insurance premiums, taxes in ports and so on. The Tonnage Laws focused on beam rather than on depth. This made tax evasions possible by deepening the hold, without increasing the breadth. These laws thus stimulated the building of deep, sluggish, flat-bottomed, flat-sided vessels (Graham, 1956: 78).

The high freight tariffs during the French Wars stimulated the rise of a new social group, professional ship owners, offering shipping services to traders and merchants. Previously, trading and shipping were both in the hands of the chartered trading companies. Trade was their main business and shipping was an instrument. By the end of the eighteenth century, shipping began to differentiate itself from trade with the emergence of professional ship owners. This process was greatly stimulated by higher freight tariffs and by the emergence of two other social groups: insurance companies and shipbrokers. Insurance companies offered a new way to deal with the risks of long-distance trading trips. Shipbrokers mediated between demand and supply of shipping services in foreign countries, which increased the efficiency in shipping, since captains would spend less time in ports looking for cargo (Broeze, 1977).

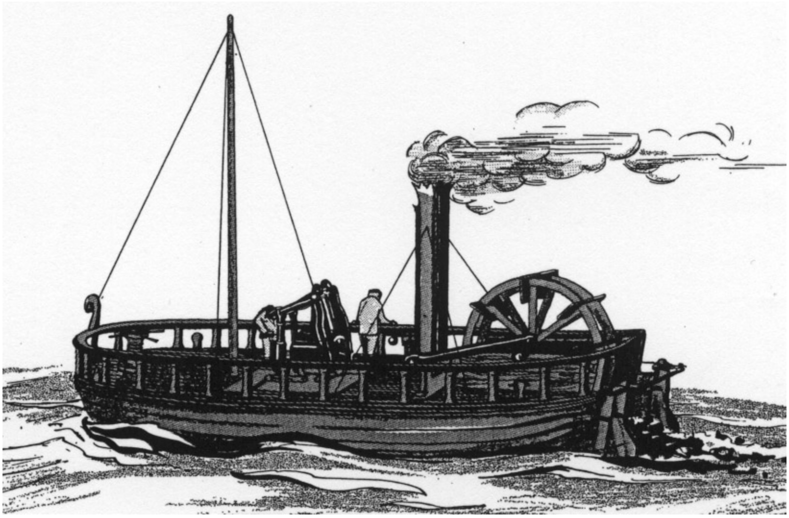
The rise of professional ship owners, insurance companies and shipbrokers was part of a wider transformation in the social network of trade and commerce. Goods were moved through increasingly denser networks of specialised middlemen, for example, factors, financiers, brokers, advertisers, wholesalers, exporters and manufacturing agents. An infrastructure for the circulation of information was created through the adoption of a series of innovations in communication, for example, journals of prices (1795), commercial newspapers (c. 1815), mercantile libraries (1820), trade journals (1831), ship-to-shore semaphore systems (1830s), agencies for advertising (1841) and credit report books distributed by subscription (1844) (Beniger, 1986: 200). New commercial institutions were developed, such as formal exchanges to conduct market transactions, concepts and usages of commercial law, and more sophisticated instruments of credit. These developments established the social network for the movement of goods on a worldwide scale.

The sailing ship regime was hindered by several problems. One problem was *speed*. On inland waterways, teams of horses and mules continued to power canal boat lines, on which sustained speeds of 4 miles per hour proved rare. On the Atlantic Ocean a sailing ship could be expected to take seven weeks plus or minus a month. A second problem was the *lack of regularity and predictability*. Uncertainty about times of arrival and departure made it difficult for traders and merchants to plan transhipments and further distribution of goods. This problem created a fertile environment for innovations that provided more regularity (Broeze, 1977: 135). A third problem was the *lack of control and coordination* in long-distance trade. Mail was the main means for communication between resident merchants and their captains, and it was transported as slowly as merchant vessels. Merchants had no up-to-date information about the quantity of goods offered in ports and their prices. Resident merchants usually gave their captains a set of general instructions on how to act in varying circumstances (for example, sail to another port if there is no trade, wait for prices to drop if they are too high). If these instructions did not suffice, the captain sent a letter back home to the merchant and waited for new instructions. This meant that a response to a change in market conditions could be implemented sometimes months after the fact. Since there could be huge variations in prices over time or between different ports, merchants faced large risks.

In reaction to these problems, four American traders pioneered a major innovation in shipping in 1818. They founded the first scheduled packet service, the Black Ball Line, which ran between Liverpool and New York on a fixed departure schedule. This improved regularity. Customers willing to pay extra money for more regular services included merchants who needed to have certain shipments sent punctually, civil servants who had to fulfil special assignments, or anyone wanting to send an official postal item. The Black Ball Line not only offered more regularity, but also higher speed. The ships carried more sail and travelled faster, cutting the eastern crossing of the Atlantic from one month to 24 days and the western crossing from three months to 40 days (Maddocks, 1982). The service proved successful, and other liner companies were set up. By 1843 there were 24 packet boats operating regular services from New York, Boston, Philadelphia and Baltimore to London, Liverpool, Le Havre and Antwerp. These ships carried urgent shipments, pressing mail and passengers for whom speed was the essence. The fast Baltimore clippers came to be widely used as packet boats for transporting passengers and mail on the Atlantic Ocean, reaching their peak between 1825 and 1850 (Pollard and Robertson, 1979). These new transportation services also offered an improvement in communications. By dissociating mail transport from trading vessels, merchants had more control over the coordination of trade. Although packet boats introduced more regularity, their time of arrival was still uncertain because they still depended on winds.

Emergence of Steamboats in the Niche of Inland Waterways

In the context of the canal boom, more attention was paid to boat innovations. Some of this attention took shape in the form of steamboat experiments. The idea of applying steam engines to boats became feasible as steam engines became less bulky and more efficient (for example, James Watt's single- and double-acting engines in 1769 and 1782, respectively). But it was uncertain at the time how steam power could best be used to propel a boat. Several methods were advocated, for example, jet propulsion,



Source: Van Oosten (1972: 15).

Figure 4.4 The Charlotte Dundas *in 1802*

screw propellers, artificial ducks feet, paddle wheels, oars and poles (Pursell, 1995: 75). Paddle wheels gradually emerged as the dominant design.

In Britain, several experiments were conducted on the Forth and Clyde Canal. A famous steam tug, the *Charlotte Dundas* (1802), demonstrated its functionality by towing two 70-ton barges against strong winds (see Figure 4.4). The canal company, however, stopped the experiment for fear that the wash from the steamboat would damage the canal's banks (Dirkzwager, 1993). It was envisaged that the steam tug would pull sailing ships through channels or manoeuvre them in ports. France and America also conducted steamboat experiments on rivers and in ports. Speeds of between 6 and 8 miles per hour were reached, almost twice as fast as horse-drawn boats. But although technical success had been proved, commercialisation was difficult because of high costs.

The first market niche for steamboats was created by Robert Fulton on the Hudson River in 1807. His *Clermont* made an average speed of 5 miles per hour (against the flow of the Hudson), and was initially used for passenger services. The Hudson River was well suited as a market niche for the primitive steamboat. The winds on the river were poor, and high and wooded banks reduced the airflow. The Hudson flowed between large centres of commerce and the area between them was hilly and the roads were bad. It had no towpaths like many European rivers, which excluded horse-drawn boats. The route was 150 miles long and very straight, without rapids. The current was not too strong, making upstream travelling possible (Gilfillan, 1935). Furthermore, public authorities in America had a positive attitude towards faster transport modes on water. This positive attitude was related to the poor state of the American road infrastructure and to the process of westward settlement.

After Fulton's success, the steamboat diffused to other inland waterways, for example, Lake Champion, and the Delaware and Mississippi rivers. The shape of steamboats was adjusted to match local circumstances, resulting in a great variety of American steamboats (Hunter, 1949). The steamboat diffused widely in America, because it linked up with the westward expansion of settlers. Waterways were the most important infrastructure to maintain contact with the settlers, because road infrastructures were of low quality or non-existent. Between 1820 and 1860, the steamboat played a crucial role in the economy, agriculture and commerce of the United States.

The steamboat was reintroduced into Britain by Henry Bell in 1812, using his *Comet* to offer commercial passenger services on the river Clyde. In order to limit the use of coal, the steamer used sail whenever possible, resulting in a hybrid mixture. Following Bell's success, steamboats were introduced on canals and rivers. As tug boats they were used in harbours, ports and estuaries to help manoeuvre large sailing ships. The next step was from estuaries to coastal routes and crossing small seas, for which the Irish Sea, the North Sea and the English Channel provided natural opportunities (Broeze, 1982). These early steamboats were relatively small vessels, because of the limited power of the steam engines. The space for the power plant and its coal supply greatly reduced the capacity to transport freight. Thus, steamers could only exist commercially in places where there was large-scale passenger and mail traffic, supplemented by special low-volume high-value cargoes. A gradual improvement in the efficiency of marine engines allowed steamboats to be employed on routes of increasing length. Table 4.2 gives some of the major historical benchmarks of regular commercial services.

Steamboats also entered the British Navy, not as warships, but for additional functions such as towboats and mail transport. They were also used as anti-pirate steamships in the colonies (Dirkzwager, 1993). Between 1820 and 1840, about 70 steam vessels were used in the British Navy fleet, an indication of their limited importance (Fletcher, 1910).

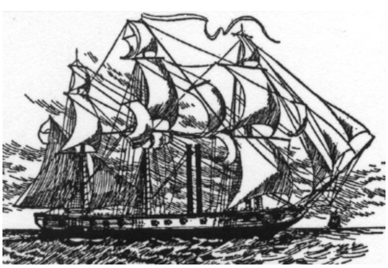
Other European countries were slow in the use of steamboats. The scope of useful functions for steamboats on their inland waterways was more limited than in America and Britain. Steamboat services faced tough competition from European road infrastructures, which were better developed than in America (Broeze, 1982).

Table 4.2 Early diffusion of regular seagoing steam navigation

Year	Route
1818	Glasgow–Belfast; Venice–Trieste
1819	Glasgow–Liverpool; Holyhead–Dublin
1821	Liverpool–Dublin; Bristol–Cork; Dover–Calais (across the Channel)
1824	Naples–Genoa; foundation of the General Steam Navigation Co.; London–Hamburg; London–Rotterdam; London–Antwerp; London–France
1826	Batavia–Surabaya
1830	Bombay–Suez (East India Company)
1835	London–Lisbon–Malaga (Peninsular Steam Navigation Co.)
1838	Bristol/Liverpool/London–New York (transatlantic)

Source: Broeze (1982: 79).

America made little progress with ocean-going steam vessels, partly because there were few productive islands to steam to (Ville, 1990). Hence, Britain pioneered the use of steamboats on the open seas (for example, the Irish Sea, the North Sea and the English Channel). But the use of steamers on oceans was thought to be impossible (Dirkzwager, 1993: 73). Although there were some isolated experiments with oceanic steamers, these were not heralded as the beginning of a new age. The *Savannah* (320 tons) was the first steamship to make the Atlantic crossing in 1819. One of the earliest steamers to cross the Atlantic in a westbound direction was a little vessel called the *Rising Star*, 1822 (Figure 4.5). The next crossing by steam came in 1833 by the *Royal William* (830 tons). In these early ships, steam engines were used as an auxiliary add-on device.



Source: Fletcher (1910: 130); courtesy of Sidgwick & Jackson.

Figure 4.5 The Rising Star *of 1822*

4.3 INNOVATIONS IN OCEANIC SHIPPING AND THE EMERGENCE OF STEAMSHIPS (1830–1848)

Landscape Developments

An important landscape development was the political and economic liberalisation in Britain between 1830 and 1850, the 'Age of Reform'. Liberal reforms also influenced formal shipping rules such as the Navigation Laws, which were relaxed during the 1830s and eventually repealed in 1849. An important step in the decline of protectionism was the termination of the East India Company's monopoly in 1834 (Graham, 1956). Starting in the mid-1830s, markets for trade and shipping increased substantially.

Industrialisation was another important macro development, initially driven by the textile industry. Huge amounts of cotton were imported and textile products exported. Britain gradually became the workshop of the world, importing raw materials and selling manufactured goods, coal, ships and financial services to the rest of the world. Industrialisation transformed the input–output flows of western economies. Metals and ore (for example, iron, copper), wool, cotton, guano and rubber were imported. Manufactured goods, textiles and coal were exported. The market for luxury products such as tea, coffee and sugar also expanded.

On land, a new transport infrastructure emerged in the 1830s and 1840s: the railway. Between 1842 and 1850 the railway network grew from 2570 to 9790 kilometres (Ransom, 1995: 567). Enthusiasm from investors led to the railway mania of 1845. Many of the planned railways were paper schemes, which vanished with the collapse of the mania in 1846. Nevertheless, a new infrastructure emerged, which enabled cheaper and faster land transport. It speeded up the 'metabolism' of society. The higher speed of land transport also increased the pressure on sea transport for higher speeds and greater predictability.

Dynamics in the Sailing Ship Regime

With the expansion of trade, the demand for shipping services picked up in the 1830s and 1840s. But the growth in trade also worsened the problems in the sailing ship regime. Professional ship owners increasingly saw the lack of control and coordination as problematic. Because mail transport was slow, market information in foreign ports was not always publicly available, and if it was, it could not quickly be sent to the resident merchant back home. Complaints were made about the quality and speed of communication (Kaukiainen, 1998).

In the 1830s and 1840s, mail steamers began to offer solutions to this problem in trade and shipping. In 1838 the British government began to stimulate the use of steamships by paying mail subsidies to shipping companies. The crucial point was that transport of mail was dissociated from the slower freight ships. Although the price of mail transport increased, so did its speed, which was much more important. Already by the mid-1840s mail was arriving in Batavia by steamer, via the Mediterranean and the Red Sea, faster than by the Cape. The London–Bombay mail had taken an average of 108 days in the East Indiamen sailing ships between 1824 and 1832. By 1840, mail was carried to India by rail, steamship and overland in Egypt in 39 days (Harley, 1985). With mail steamers, communications and coordination in the trade were greatly improved. The adoption of the new technology thus actually improved the functioning of the existing regime. Steamers had a symbiotic relationship with sail, rather than a competitive one.

The transport of passengers also expanded, to Australia and New Zealand in the 1820s and 1830s and to America in the 1840s. Packet boats flourished in this expanding passenger market, reaching their peak between 1825 and 1850.

The growth of luxury markets (for example, in opium and tea) stimulated technological development. In the high-value freight markets new ship designs emerged, particularly as the monopoly of the British East India Company was broken in 1834. Because of the high value of its trade, China was an attractive market. New American ships, the opium clippers, quickly explored this new market and were widely used from 1830 to 1850 (Pollard and Robertson, 1979). These ships were larger than the Baltimore clippers and more designed for cargo, but nevertheless fast and manoeuvrable.

In the 1840s, large, long, flat-bottomed cargo carriers emerged: the clipper ships. With their fine lines, sharp bow, broad beam and tall raking mast, tea clippers were larger than earlier ships, and reached speeds well above any previously obtained. High speed was an important criterion in the Far Eastern tea trade, because the quality of tea declined during transportation (Harley, 1971: 224). Many 'extreme' clipper ships were built, with very sharp bows, sacrificing cargo capacity for speed. The most important era of the tea clippers extended from 1840 to 1860 (Pollard and Robertson, 1979). The new clipper ships were developed by American shipbuilders. British shipbuilding was less innovative. The British Tonnage Laws were altered in 1836 to make tax evasions more difficult, but the new laws were not made compulsory until 1855, holding back improvements in British ship design. Stimulated by their innovative ship design, American shipping expanded more rapidly than the British fleet (see Table 4.3).

By 1850, however, American shipbuilders began to experience construction problems. To be fast, ships had to be long. But as ships increased in

Table 4.3 Registration of ship tonnage in United States and Britain, in '000 tons

	United States	British Empire
1815	1368	2681
1825	1423	2554
1835	1825	2784
1845	2417	3714
1850	3527	4233
1854	4803	5166
1858	5050	5610
1861	5488	4895

Source: Harrison (1990: 36).

size, they met problems of longitudinal strength. Wooden ships were approaching their 'natural limit' of 275–300 feet. It was becoming more and more difficult to obtain timber that was long enough, so builders had to construct beams out of smaller pieces, resulting in less strong ships. In the early 1850s, American prestige was under threat (Calhoun, 1973). British shipbuilders were able to deal with this problem by taking advantage of their knowledge and experience in working with iron. They gradually moved over to partial metal construction in the 1850s, which helped them to regain the upper hand against American shipbuilders.

Further Development of Steamships in the Mail Transport Niche

Steamships could ply the seas, because they linked up with the problem of long-distance communication and coordination. In 1838, the British government decided to pay mail subsidies to steamship companies on particular routes in the British Empire. Mail steamers could deliver mail faster than sailing ships, which was beneficial to public servants and private merchants. Two steamship companies were awarded the mail transport franchise to America, the Cunard Line and the Great Western Steamship Company. The steamships were operated as line services, having fixed departure *and* arrival times. Steamships thus also linked up with another problem in the sailing ship regime, unpredictability and irregularity. Line services, in effect, introduced new functionalities in oceanic shipping, greatly improving regularity and reliability. The steamship companies carried first-class passengers and mail, and could not have existed without mail subsidies. Between 1838 and 1862, a global network of British intercontinental steam companies was created on the basis of imperial mail subsidies (see Table 4.4).

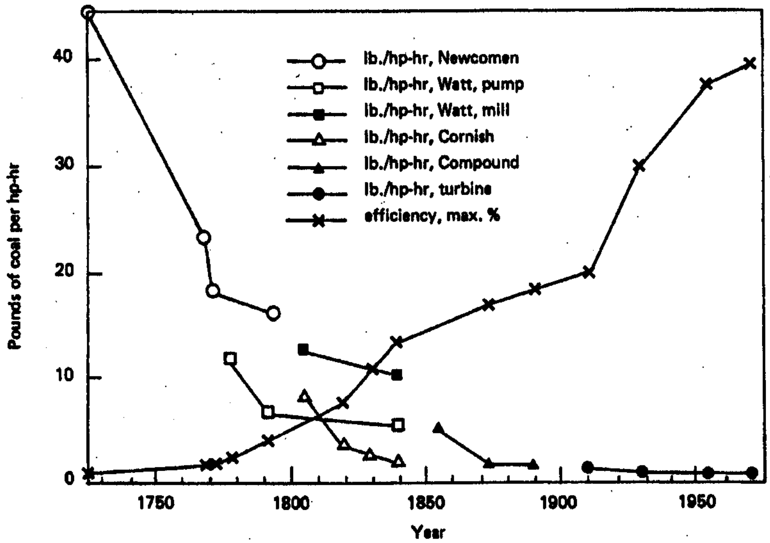
Table 4.4 Steamship lines created with British mail subsidies

Year	Route	Company
1837	Portugal, Spain	Peninsular Steam Navigation Company
1839	Liverpool–Halifax–Boston (1848: New York)	Cunard Line
1839	Southampton–West Indies	Royal Mail Steam Packet Co.
1840	Britain–Egypt–India	Peninsular & Oriental Steam Navigation Company (P&O)
1844	India–Singapore–China	P&O
1845	Valparaiso–Callao–Panama	Pacific S.N. Co. (founded in 1840)
1850	Britain–Brazil–La Plata	Royal Mail
1852	Britain–West Africa	African Steamship Co.
1852	Singapore–Australia (1859 from Ceylon)	P&O
1856	Britain–Canada	Allan Line
1857	Britain–South Africa	Union Steamship Co.

Source: Broeze (1982: 83).

The operation of oceanic steamships was not without its problems, however. First, net carrying capacity was reduced: steam engines had a high coal consumption and vast quantities of coal had to be carried on board. Second, paddle wheels became submerged or rose out of the water in rough seas, which damaged the engines, reduced the functioning of the paddle wheels, and diminished the ship's manoeuvrability. Third, the heavy weight of the boilers, condensers and steam engines caused the wooden hull to bend and stretch. As steam engines and boilers became more powerful and heavier, this problem was exacerbated.

These three problems were picked up in the 1840s by an emerging steamship community. The mail subsidies and expansion of market niches for steamers provided 'space' for innovations. Shipbuilders, engine makers and inventors tried to find solutions for the issues on the problem agenda. To deal with the first problem, the coal efficiency of steam engines had to be improved. The main heuristic for achieving this was to raise boiler pressure, which was done by making boilers heavier and more resilient, using stronger metals. Steam pressure in marine boilers was raised from 5 psi (pounds per square inch) in the 1830s to 10 psi in the 1840s and 20 psi in the 1850s (Graham, 1956). Another search direction related to the design of boilers. Between 1840 and 1850, tubular boilers were generally adopted, being lighter and more compact (Fletcher, 1910). The efficiency of steam engines was also improved by better lubricants, which reduced friction



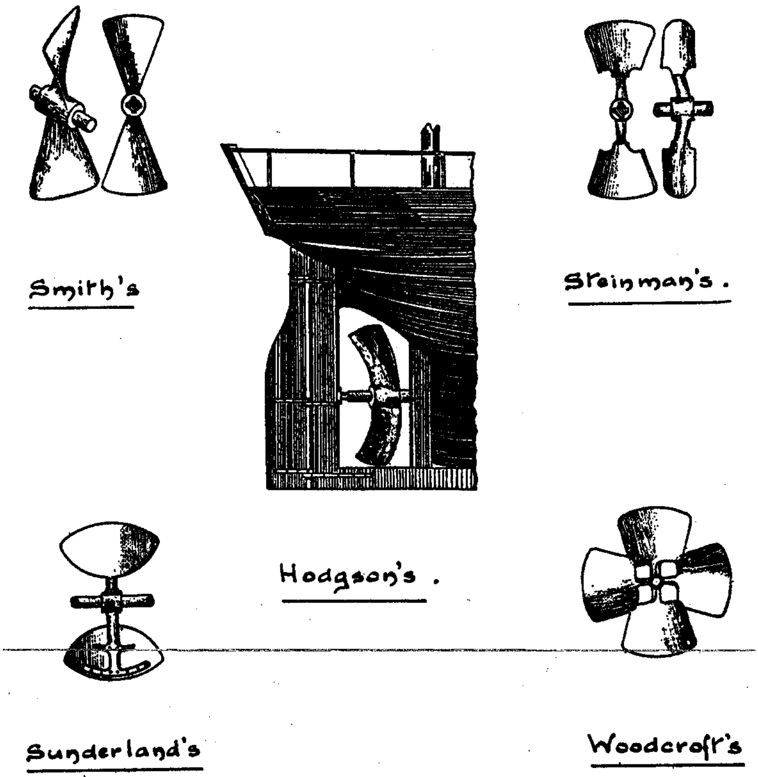
Source: Ayres (1989: 16); courtesy of the International Institute for Applied Systems Analysis.

Figure 4.6 Fuel consumption and thermal efficiency of steam engines

between moving parts. As a result of these innovations, fuel consumption and thermal efficiency gradually increased (see Figure 4.6).

The problems with paddle wheels gave rise to several search directions. Incremental changes were tried with regard to the position of the paddle wheels, for example, on the side of the ship, or at the back. There were also pioneers who investigated an alternative propulsion system: screw propulsion. Two early inventors, John Ericsson and Sir Francis Pettit Smith, both acquired patents in 1836 (Gilfillan, 1935). In 1836 and 1837 they aroused great interest with five screw vessels for demonstration purposes. Although the basic principle of screw propulsion was demonstrated, practical problems needed to be solved. There were no clear design rules to calculate the best shape of screw propellers and consequently many shapes were tried out in practice (see Figure 4.7).

Because screw propulsion required a higher number of revolutions than paddle wheels, new devices were needed to drive the screw fast enough. Ericsson proposed to connect the screw and engine with a screw shaft, using toothed gearing to achieve high speeds. Another problem was how to connect the screw to the steam engine and how to place it in the ship: since the screw was placed underwater, it required a watertight hole for the



Source: Smith (1937: 76); courtesy of Cambridge University Press.

Figure 4.7 Various screw propellers tried in the Rattler, *1845*

screw shaft (Dirkzwager, 1993: 79). The design of an underwater stern bearing was difficult, and the stern gland fittings had long been a source of great anxiety (Smith, 1937: 78). Another major issue was the vibration problem. The higher number of revolutions in screw propulsion led to vibrations and more stress on the ship's hull. Wooden ships in particular ran the risk of being shaken apart, which stimulated a gradual shift towards iron hulls.

In sum, many problems had to be worked out to make the application of screw propulsion practicable. In the late 1830s, inventors tried to sell their innovative ideas to the British Navy. Screw propulsion had the advantage that the machinery and the screw could be placed below the waterline and out of reach of shot, leaving the decks clear for armaments, something more difficult with paddle wheels. But the Navy was reluctant to fund the innovation, preferring that others footed the bill (Lambert, 1999). However, the mercantile community proved more receptive, and screw propulsion caught the attention of Isambard Kingdom Brunel, an engineer working on the iron vessel *Great Britain*. Brunel adopted it for the *Great Britain* (1843), which became the first screw-propelled steamer to cross the Atlantic. Many ship owners, however, thought that screw propulsion was experimental and too risky. Before 1850 there were still problems with machinery, stern bearings and watertight glands (Lambert, 1999). Hence, ship owners were slow to adopt screw propulsion.

The third innovative trajectory was iron hulls. Some innovative shipbuilders began experimenting with iron ships in the 1830s. This was not easy, because iron ships required new skills and competencies, for example, riveting iron plates together (Wimmers, 1998). Rivet-working skills were very different from woodworking skills. Furthermore, iron required different process techniques. It had to be heated, hammered, flattened, punched, cut and welded. Shipbuilders used to working with wood possessed neither the required skills nor the machines to build with iron. Hence, the construction of iron ships depended on outsiders, ironworkers and boilermakers. Iron hulls and steam engines required different resources, such as coal and iron, skilled engineers and a supply of cheap workers. Builders of iron steamships moved from London, the heart of wooden shipbuilding, to northeast Britain where other heavy industries were located. The first iron ships encountered great scepticism – some people believed that iron ships would sink because iron was heavier than water (Fletcher, 1910). For shipbuilders the use of iron as a new material created new problems, because many of the traditional building rules and practices were no longer applicable. Scott Russell, a shipbuilder and vice-president of the Institution of Civil Engineers, wrote in 1875:

> '[To the generation before] it would have been incredible . . . to build ships of a material that will not float, that is eight times heavier than water, and that quickly sinks to the bottom of the sea. . . . Few of the rules and traditions of that craft, which showed how to cut skilfully the members of a wooden hull out of the gigantic oaks of a forest, have survived or can be of any help. Neither the shape, the proportions, the joints, the fastening, nor the functions of the various members forming the hull of an iron ship, can be said to resemble those of wood.' (cited in Garrat et al., 1973: 71)

Shipbuilders initially tried to build iron ships on the basis of rules and criteria from wooden shipbuilding practices, but this was not successful. The early iron ships often suffered from instability:

> 'Ships were launched in my time, so ill calculated in quality, that their first evolution, on reaching the water, was to turn upside down and to stay that way. . . .

> We remember another fleet of iron ships, built for another great company, being all actually completed before their utter unseaworthiness was discovered; and we remember the cure to have been the building of a great brick wall, in cement, on the inside of the iron hull, on the bottom, so that the weight of the bricks should keep the bottom of the ship from turning upside downs in a sea-way, and for many years these ships were kept on end by these means alone.' (Russell cited in ibid.: 69–70)

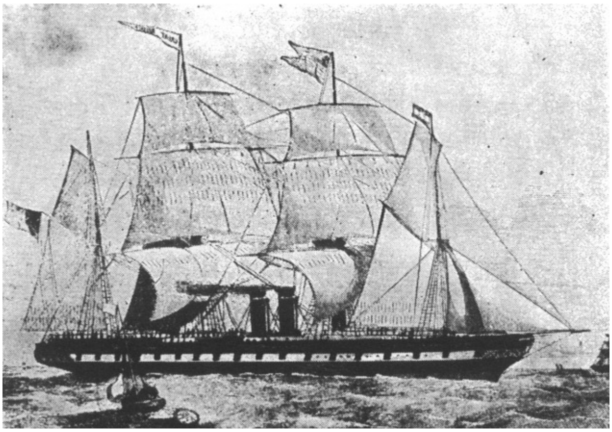
New design rules were articulated through trial and error. Gradually it was discovered that iron had some advantages over wood: iron withstood fire and the vibration from steam engines and screw propellers. Because of its greater strength the hull could be thinner, even to the point of being lighter than wooden ships (Gilfillan, 1935). Official recognition of iron ships dates from 1837 when the steamer *Sirius* was classified in Lloyd's Register. High insurance premiums had to be paid for early iron ships. No rules for the construction of iron ships were issued by Lloyd's until 1855 (Smith, 1937).

Iron ships were first used on inland waterways and short-distance routes. A major problem for their use on oceans was that iron hulls disturbed the compass (ibid.: 96). It was not until 1855 that a proper compensating arrangement for the compass was developed, enabling its confident use on an iron ship (Dirkzwager, 1993). Another problem was the rapid fouling of the hulls by marine growths, which slowed down the ship. The Navy was initially sceptical about iron hulls, after shooting tests in the 1840s showed that iron hulls splintered and fragmented when hit by bullets, causing major damage to the inside of the ship (Dirkzwager, 1993). It was the mercantile community that sponsored the first experimental projects with iron hulls on oceanic ships. Hence, the use of iron by Brunel to build the *Great Britain* (1843) was a major experiment, as was the use of screw propulsion. Although the *Great Britain* successfully crossed the Atlantic to New York, it would take further demonstrations before iron ships were widely accepted. The *Great Britain* was a hybrid form between sailing ships and steamships (see Figure 4.8).

4.4 THE DOMINANCE OF SAILING CLIPPERS AND THE TAKE-OFF OF STEAMSHIPS IN PASSENGER TRANSPORT (1848–1869)

Landscape Developments

Liberalisation and industrialisation were two macro developments, which continued to stimulate world trade and shipping. Between 1840 and 1887 there was a sevenfold increase in seaborne commerce (see Table 4.5). As part of the liberalisation process, the British Navigation Acts were repealed in



Source: Gilfillan (1935: 152); courtesy of Follett Publishing Company.

Figure 4.8 The Great Britain*, 1843*

Table 4.5 Merchandise carried by sea, 1840 and 1887, in '000 tons

Commodity	1840	1887	Commodity	1840	1887
Coal	1,400	49,300	Jute	–	600
Iron	1,100	11,800	Meat	–	700
Timber	4,100	12,100	Coffee	200	600
Grain	1,900	19,200	Wine	200	1,400
Sugar	700	4,400	Salt	800	1,300
Petroleum	–	2,700	Sundries	9,180	33,750
Cotton	400	1,800			
Wool	20	350	Total	20,000	1,40,000

Source: Craig (1980: 18).

1849. The British example was followed by other European countries, but much more slowly. Sailing ships carried much of this growing trade. Steamships later linked up with the growth in world trade, and reinforced it.

Another landscape development was mass emigration from Europe. Between 1820 and 1920, about 60 million Europeans left for the New World. Some 60 per cent of them went to America, others to Australia, New Zealand, Canada and South America (O'Rourke and Williamson, 1999). The first wave of mass emigration occurred in the late 1840s. Between 1846 and 1855 more than 2 million Europeans left for America (Maddocks, 1982). This mass migration was caused by several other macro developments. One development was the Irish potato famine (1845–49), related to failing potato crops in successive years. A second development was the European political revolution of 1848: many people emigrated to escape religious or political persecution. Third, wages in America were higher than in Europe: many people wanting to escape European poverty were attracted by high American wages, and this was stimulated by the gold rushes in California (1848) and Australia (1851). Most of the European emigrants were poor and travelled in sailing packets. The rich minority, travelling in luxury cabins, paid about ten times as much as the poor on sailing packets and even more on steamships. This growing passenger market provided a window of opportunity for steamships to acquire wider market niches.

Dynamics in the Sailing Ship Regime

Shipping in the 1850s and 1860s was characterised by strong growth in passenger and freight transport. The growing markets for shipping services produced long-term optimism in the sector, known as 'the Golden Fifties' (Palmer, 1985). In these prosperous times, there was money for innovation. British shipbuilding enjoyed a new vitality. Stimulated by the new Tonnage Laws, which became compulsory in 1855, British shipbuilders followed the American example and also began building clipper ships. An important difference was that British shipbuilders gradually shifted from wood to iron as building material for sailing ships. This was partly a response to problems of longitudinal strength. Another reason for the shift towards iron was the increasing scarcity and rising price of timber in Britain (Harrison, 1990). Because of production improvements, iron actually became cheaper than wood as a building material in the 1860s. In America, however, timber remained considerably cheaper than iron. The initial advantage of the Americans gradually turned into a disadvantage as the technical frontier moved on.

Iron entered British shipbuilding in a gradual and stepwise process. Iron was first used as an add-on, to strengthen the existing wooden constructions in the form of knees or connections between the deckhouses and the ribs. As a second step, composite sailing clippers were built in the 1850s, consisting of an iron frame and wooden planking. This was a hybrid, intermediate form between iron and wood. In the late 1850s and early 1860s, ships with all-iron hulls and steel masts became more common. This third step was stimulated by a substantial revision in 1863 of Lloyd's Rules, lowering the insurance premiums of iron ships (Harley, 1973).

The gradual shift towards iron was an important factor in the re-emergence of British shipbuilding. Another major factor was the American

Civil War (1861–65), in which about 40 per cent of American seagoing tonnage was lost (Harrison, 1990).

Take-off of Steamships in the Passenger Transport Market Niche

The percentage of steamships in British registered tonnage grew from 5 per cent in 1850 to 10 per cent in 1860 to 17 per cent in 1869. Hence, the 1848–69 period can be characterised as the take-off phase in the diffusion of steamships. This rapid growth was possible, because steamships linked up with the growing market of transatlantic passenger transport, stimulated by European emigration. Rich passengers increasingly opted for steamships, which offered higher speed, greater regularity and more comfort. They were willing to pay extra for this kind of transport. Early steamship companies took little interest in poor emigrants, leaving them to travel in the sailing packets. But as steamships grew in size, they also took on board poor emigrants, stowing them in second- and third-class decks.

Stimulated by growing passenger markets, liner companies turned into large corporate and professional companies, operating fleets of liners which ran to a regular timetable of departure *and* arrival times (Ville, 1990). A major organisational change was the introduction of a new kind of ownership: the joint-stock company. This new institution made it easier to acquire capital to buy steamships, which were more capital intensive than sailing ships. When steamships were operated as a fleet of liners, the large companies were set up as joint-stock companies.

The rise of liner companies also contributed to a change in management practices. Managers had to pay more attention to matters such as fleet management, financial accounting, cost control, detailed budgeting and long-range planning (Sloan, 1998). Financial accounting tried to deal systematically with matters such as initial cost, useful operating life, replacement cost, vessel depreciation and residual value. The shift to new management practices took place through trial and error. Early steamship owners did not always appreciate the new capital aspects related to steamships, and sometimes continued to rely on conventional financial standards from the sailing ship enterprise. In particular, they often failed to grasp the matter of steam vessel depreciation (ibid.). The management of shipping companies became a specialised occupation for trained professionals. Ownership and management were increasingly separated.

Because of its profitability, the Atlantic Ocean became a competitive arena for liner companies. To distinguish themselves, liner companies ordered ships that were ever larger, faster, safer, more luxurious and modern. The importance of speed was emphasised, with the institution of the ‘Blue Riband’ prize in the 1860s for the fastest crossing of the

Atlantic. This resulted in an innovation race, which provided 'space' for further work on new technical elements such as iron hulls, screw propulsion and better steam engines. Developments in these three elements are discussed below.

By the mid-1850s, iron became more accepted as a building material. The Navy changed its tune after grenades were introduced during the Crimean War (1853–56), since these weapons caused substantial internal damage to wooden ships. In response to grenades, the French developed the frigate *La Gloire*, using heavy iron plates as armour. The British Navy decided to go one step further and adopt iron hulls in 1859 (Dirkzwager, 1978). The mercantile community also developed more confidence in iron hulls, as was indicated by Lloyd's lowering of insurance premiums of iron ships in 1863 (Harley, 1973). The wider acceptance of iron hulls required complementary innovations, for example, the development of anti-fouling paint to prevent marine growth. Initially a composite wood and copper sheathing was wrapped around iron vessels. Another solution was to develop coating composites and anti-fouling paints: salts of mercury, lead, antimony, zinc and copper were tried. In 1860, Captain Rathjen achieved worldwide success with his quick-drying alcoholic shellac solution (Gilfillan, 1935). As more iron ships were built, shipbuilders began to move beyond the experimental stage. They gradually learned that iron allowed for new functionalities and designs. Early shipbuilders saw iron as a simple substitute for wood, and they translated wooden construction principles into iron. The iron keel, for example, was built in the same way as a wooden keel would be. Only later was its shape and method of construction changed (Dirkzwager, 1993). During the 1850s new designs were tried, for example, the double hull (which was used both as a safety compartment and as a set of containers for water ballast or oil fuel), watertight bulkheads (which were used for safety and strength), and bilge keels (which were lateral fins outside the ship to reduce rolling) (Gilfillan, 1935). In addition, the use of iron made it possible to build much larger ships; iron was more durable than wood, lengthening the useful life of ships; and iron hulls could much better withstand the constant vibration of screw propulsion, thus solving the 'vibration problem'.

Screw propulsion gradually established itself as the dominant propulsion mode in the 1850s and 1860s. In the early 1850s, the Navy began switching to screws. Commercial steamship lines were somewhat slower in adopting the screw, because of the damage from vibration to wooden ships. However, this problem was solved as iron hulls came to be more accepted by the mid-1850s. One of the first liner companies to adopt the screw was the Inman Line in 1850. The growing confidence in screw propulsion is indicated by the lowering in the 1850s of insurance premiums on screw ships from 4 to 1.25 per cent (Lambert, 1999).

To reduce the amount of coal on board, coal efficiency of steam engines had to be further increased. A promising step forward was the compound engine, which used high-pressure steam twice to drive an engine. The steam from the first cylinder, with high initial pressure, would be passed to a second cylinder of greater bore, where there was less pressure per unit of area. Thereby the amount of power from a given amount of steam could be increased considerably. Compound engines were first developed for land-based, stationary applications such as in mines and factories. In the late 1820s and 1830s, compound engines were used to some degree on steamboats for inland waterways (Verbong and Van Overbeeke, 1994: 230). But compound engines were not used on the high seas, because injection of salt water, to condense the steam, caused problems. Salt would form sediments on the inside of the cylinder, reducing its working efficiency and creating explosion problems. A new way of steam condensing had to be developed before compound engines could be used efficiently on oceans. The answer was the surface condenser, where steam was condensed by passing it through small tubes cooled by water on the other side. Already in 1832 Brunel had obtained a patent for a surface condenser, but technical problems regarding airtight sealing prevented its application. However, in the 1850s practical surface condensers were developed, enabling the use of oceanic compound engines. The first experimental compound engine with a surface condenser was installed on an oceanic steamer in 1854 and was used successfully to cross the Atlantic (Broeze, 1982). The engine used about 3¼ lb coal per horsepower per hour, compared to 4 to 4½ for contemporary engines (Craig, 1980). The efficiency of the compound engine was hampered, however, by the poor quality of boilers, which made it impossible to sustain high working pressure. Before 1860, steam pressures were generally not higher than 30 psi. But better lubricants and the use of steel improved the functioning of boilers, enabling higher pressures of 40 psi in the early 1860s and 50 psi in the late 1860s. The compound engine was crucial for using steamships on long-distance routes, because it greatly improved coal efficiency (Graham, 1956). With the compound engine, steamships entered long-distance market niches in freight shipping, initially in the Chinese tea trade, where merchants were willing to pay a premium price for high speed. The first steamer, the *Agamemnon*, reached China in 1866 via the Cape of Good Hope. Because the compound steam engine consumed almost 40 per cent less fuel than earlier steamers, steamships were able to compete successfully with the celebrated 'tea clippers' in the tea trade (Craig, 1980).

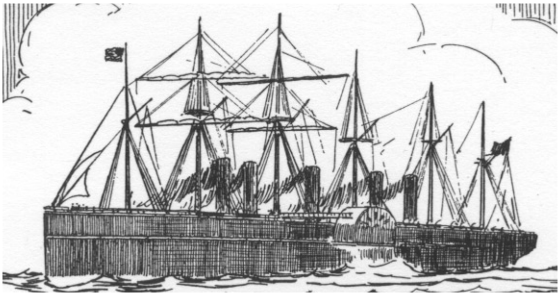
Table 4.6 Characteristics of two steamship regimes

Low pressure	High pressure
Low temperature	High temperature
Single expansion	Compound or higher expansion
Organic lubricants	Mineral lubricants
Box boiler	Cylindrical, tubular boiler
Wood and iron (composite)	Iron and steel
Little engineering design, much craft workmanship	Much engineering and metallurgy
Jet condensers	Surface condensers
Low rate of revolution	High rounds per minute
Paddle-wheels, large	Screw, twin screws or small paddle-wheels
Weak engine, slow steamer	Powerful and fast ship
Steam auxiliary or sail auxiliary	All steam power
Freshwater navigation, small seas, early ocean travel	Ocean voyaging

Source: Gilfillan (1935: 130).

During the 1850s and 1860s, the new technical elements were gradually linked together, resulting in a new technical steamship regime centred on high-pressure compound steam engines, surface condensers, iron hulls and screw propellers. The differences with the earlier steamship regime are summarised in Table 4.6.

An early ship, in which many new elements were used, was the *Great Eastern*, launched in 1858 (Figure 4.9). This ship represented a gigantic leap forward, being six times larger than other ships of its time. It was 211 metres (692 feet) long and 18,915 gross tons (Gilfillan, 1935). It was designed to be able to steam to Australia and back without re-coaling (because there was not yet a worldwide coal infrastructure). To reduce its coal consumption, the *Great Eastern* had a peculiar combination of steam and sail, which was used to assist the ship around the Cape. Although the *Great Eastern* was admired for her engineering, the coal consumption of her engines had been seriously underestimated, and she was never used for the Australia run. The ship also rolled excessively in bad weather, which reduced passenger confidence in the ship. Not enough was known at the time about the behaviour of a very large steamship on the high seas. As a commercial venture the *Great Eastern* was a complete failure. It never made any money because it was too large and too progressive, running too far ahead of its time. There were not enough passengers, cargo or docks to make her a commercial success (ibid.).



Source: Gilfillan (1935: 153); courtesy of Follett Publishing Company.

Figure 4.9 The Great Eastern*, 1858*

4.5 WIDE DIFFUSION OF STEAMSHIPS (1869–1890)

Landscape Developments

Mass emigration from Europe continued to be an important landscape development. In the 1880s and 1890s emigration speeded up again, when about 600,000 people left per year. After the turn of the century the emigrant stream increased further to about 1 million emigrants per annum (O'Rourke and Williamson, 1999: 119). Between 1892 and 1920, 12 million migrants passed through New York. Because of this landscape development, passenger transportation continued to be a growth market. Many new liner companies were formed and existing companies expanded their passenger services.

Another landscape development was the continued growth and diversification of world trade, related to industrialisation. The industrial development of western countries relied heavily on imports of cheap raw materials and exports of industrial products. Between 1850 and 1913, world trade per capita grew at over 30 per cent per decade (Ville, 1990). British ship owners carried most of this trade: as late as 1914, no less than 40 per cent of seaborne trade still touched on Britain (Palmer, 1985). The freight shipping market not only grew, but also diversified, both in terms of shipping routes and the kind of freight. Shipping routes to new continents were opened up, while established routes were expanded. Latin America became an important source of guano and metallic ore. Australia and Latin America became a major provider of meat after refrigeration was developed in the 1880s. America and Russia became major suppliers of cheap wheat. After 1880, African countries such as the Congo and Northern Rhodesia (Zambia) became important suppliers of minerals and ores. The result was a more diverse freight market, consisting of many market niches with different selection criteria. Many new freight markets consisted of low-value high-volume bulk transport (for example, coal, wool, ores, cotton, meat, minerals), for which low freight costs were more important than high speed. These new market niches provided 'space' where sailing ships could hold on for decades. The growth of world trade and steamships had positive feedbacks on each other. On the one hand, growing markets provided space for the further development of steamships. On the other hand, improvements in steamships (higher coal efficiency, economies of scale) enabled lower freight tariffs. Relative freight rates moved down from 100 in 1830 to 24 by 1910–14 (Ville, 1990). The falling transport costs further stimulated globalisation and the expansion of world trade.

Wide Diffusion of Steamships and Competition with Sail

A major change in the physical landscape was the opening of the Suez Canal in 1869, which greatly shortened distances to the east (see Table 4.7). Because the Suez Canal was unsuitable for sailing ships, on account of the few and variable winds, this gave steamships a great comparative advantage on Far Eastern trades. The result was that steamers were suddenly able to compete successfully with sailing ships in the Indian and Asian markets. The Suez Canal gave steamships their first major market in freight shipping. It led to a five-year boom in the construction of new steamers, the so-called 'steamship mania' (1869–74). During this period, the percentage of steamers in the British fleet increased from 17 per cent of total registered tonnage to 31 per cent. But the boom was followed by a backlash. The shipping depression of 1874 caused a revival of sailing ships because they had lower operating costs, and could afford, in days of bad port facilities, to wait for

Table 4.7 Effect of Suez Canal on shipping routes to East India

Route from London to:	Distance around Africa in km	Distance through Suez Canal in km	Saving in %
Bombay	19,800	11,470	42
Calcutta	21,700	14,600	33
Singapore	21,100	15,250	28
Shanghai	25,500	19,300	24

Source: Broeze (1982: 95).

a cargo (Graham, 1956: 85). Some steamers were actually converted back to sail (Ville, 1990: 54).

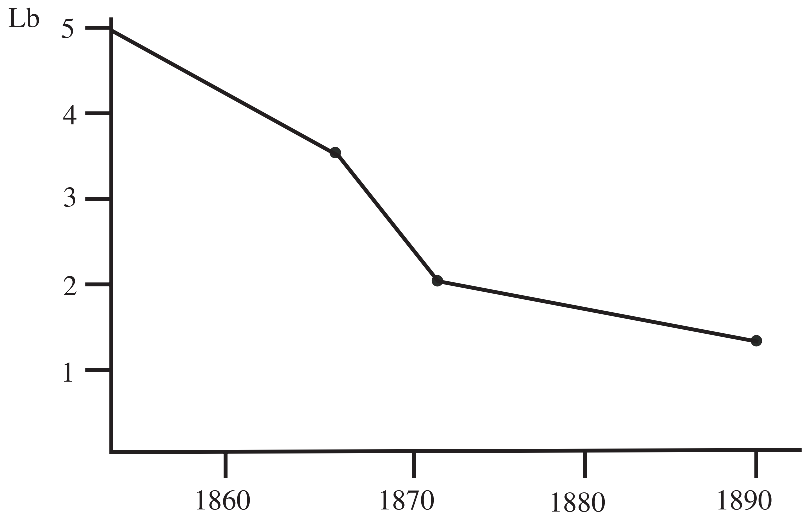
Before 1869, steamships had been used in small freight shipping market niches. Steamships could only be used cost efficiently over short distances, because they used a lot of coal. There was a certain 'distance margin' below which steamers could compete with sailing ships. Above this margin, sailing ships were generally cheaper. As steamships were improved, the distance margin increased. In the 1850s, British steamers entered some freight trades with northern Europe and the Baltic countries, and around 1865 they began competing in the Mediterranean fruit trade. With the introduction of the compound steam engine in the mid-1860s, steamships entered the Chinese tea trade. By the end of the 1860s, the distance margin between sail and steam was raised to 3000–3500 miles, enabling the use of freight steamers on the North Atlantic grain trade (Harley, 1988). The Suez Canal opened up the large market niches of the Indian and Chinese trades. As coal efficiency improved (higher boiler pressure, compound engine), steamships came to be employed on ever-longer routes. Table 4.8 gives an overview of routes and dates on which steam became competitive with sail. The diffusion of steamships in freight shipping was gradual for three complementary reasons: (1) gradual technical change and performance improvement of steamships; (2) defence by sailing ships; and (3) changes on wider dimensions of the socio-technical regime, for example, ports, coal infrastructure, shipyards.

(1) Improvements in steamships consisted of a continuous accumulation of larger and smaller innovations. The decrease in coal consumption per

Table 4.8 Freight routes and dates when steam became competitive with sail

Date	Voyage (route)	Distance (miles)
1855	Northern Europe	500
1865	Mediterranean fruit and cotton; Chinese tea trade	Up to 3000
1870	North Atlantic grain trade; Bombay via Canal	3000 6200 via Suez Canal
1875	New Orleans cotton	5000
1880	Calcutta	8200 via Suez Canal; 11,500 via Cape
1895	West Coast of America, grain, ore	13,500 to San Francisco

Source: Harley (1985: 177).



Source: Harley (1985: 176).

Figure 4.10 Coal consumption of marine steam engines (lb per indicated hp/hr)

indicated horsepower per hour went down gradually over the years, increasing competitiveness (see Figure 4.10).

A range of innovations made the increase in coal efficiency possible. Boilers were gradually improved, facilitating higher steam pressure. Higher pressure depended critically on better quality steel, as well as better lubricants which improved airtight sealing (Gilfillan, 1935). Innovations such as super-heaters and forced draught (which allowed the use of poorer quality coal) also enhanced coal efficiency. Based on a better theoretical understanding of thermodynamics, the compound engine was further improved, eventually resulting in the triple-expansion engine in the 1880s. Table 4.9 indicates the resulting improvements.

Other innovations were also involved in lowering operating costs of steamships. As ships became larger, relative transport costs per ton decreased (economies of scale). Lower coal consumption meant fewer stokers and reduced labour costs. The price of iron decreased rapidly from the mid-1870s. From the early 1870s to the mid-1890s, average building costs of steamer hulls more than halved (Kaukiainen, 1992). Because steamships were capital intensive, rapid turnarounds in ports were important. Improved port facilities and cargo handling (for example, cranes), reduced the time

Table 4.9 Improvements in the performance of steam engines

	1872	1881	1891	1901
Boiler pressure (pounds per square inch)	52.4	77.4	158.5	197
Revolutions per minute	55.67	59.76	63.75	87
Piston speed (ft per min)	376	467	529	654
Coal consumption (per hp per hour, lb)	2.110	1.828	1.522	1.48

Sources: Craig (1980: 14).

spent in ports, improving useful days per year. The continuous accumulation of innovations gradually improved the steamship's competitiveness, stimulating its diffusion.

(2) A second reason for the gradual diffusion of steamships was the defence by sailing ships. The first defence strategy was the technological innovations in sailing ships themselves, leading to increased cargo capacity, higher speed and less crew. Cargo capacity was increased by building larger sailing ships making use of the possibilities of iron hulls. By 1870, iron sailing ships had double the space for cargo in proportion to tonnage (Graham, 1956). Higher speed was achieved by new hull designs, longer ships and additional masts and sails. More sail increased the speed of sailing ships but reduced the ship's manoeuvrability. The trend to use more masts and sail culminated in the *Thomas W. Lawson*, a schooner with seven masts and much sail, built in 1902. The ship was very fast, but had low manoeuvrability. In fact, the ship was so unstable that it capsized while at anchor during a severe gale in 1907 (Foster, 1986). To reduce crew costs, labour-saving machinery (for example, for rigging) was introduced. With these new machines, sailing ships could be manned by about 30 per cent the number of men (Graham, 1956). The renewed sailing clippers, with their iron hull, steel masts and rigging and labour-saving machinery, were strong competition for steamships in the ocean trades of the 1870s and 1880s. The improved sailing ship held its place as the principal overseas carrier well into the 1880s. Such improvements exemplified the general pattern whereby technology is improved when it is challenged by a new technology. Because this pattern is so obvious with sailing ships and steamships, it is called the sailing ship effect.[1]

The second defence strategy was that other innovations, not in the ship itself, improved the functioning of sailing ships. Steam tugs, for example, helped large sailing vessels manoeuvre in ports. This encouraged the building of sailing ships of even greater cargo capacity, as manoeuvrability in ports was no longer a problem (ibid.). Steamers thus not only competed with sailing ships, but also helped them. Another innovation was the growing knowledge of oceanography and the creation of reliable charts of winds and currents. Matthew Maury's standard work of 1850 helped captains find the best route in any particular month or week. As a result, the average passage to the equator was shortened in the early 1850s by 10 days. The journey from Britain to Australia was shortened from 125 to 92 days. This enabled fast clipper ships to compete with steamships (ibid.).

A third defence strategy of sailing ships was to evade to new markets, as steamships threatened them. This strategy was made possible by the expansion and diversification of world trade. Many of the new markets were bulk cargoes, consisting of raw materials. Because low costs were a more important selection criterion in bulk markets than high speed, sailing ships found good employment there (ibid.). A major bulk market for sailing ships was the transportation of coal from Britain to build up a worldwide infrastructure of coaling bunkers for steamers (along the African coast, Hong Kong, Singapore and the Latin American coast). This is another example of the symbiotic relationship between steamships and sailing ships.

(3) The third reason for the gradual diffusion of steamships was that many elements of the socio-technical shipping regime had to be adapted, something which inevitably took time. Ports and harbours, for instance, had to be adapted. As the size of steamships increased, longer, wider and deeper locks were needed (Pearsall, 1996). Large steamships called forth a new generation of port facilities, tailored to the need for a rapid turnaround. Docks were fitted with modern cargo handling and unloading gear, for example, cranes and conveyor systems (Jarvis, 1998). The introduction of bulk cargo-handling machinery was sometimes the object of controversy and social struggle because labourers feared that they would be made redundant.[2] A range of new transhipment facilities emerged to store cargo that waited for further transportation. All these changes in ports required huge investments (Jackson, 1998).

Another change in the socio-technical regime was the creation of a worldwide infrastructure of coaling stations. This fuel infrastructure was created gradually in a stepwise fashion. The worldwide transportation of coal provided a large market niche for sailing ships.

Third, shipyards were also substantially transformed. Geographical location, job skills and workplace organisation were all altered as shipbuilding moved from its craft skill tradition to a heavy engineering industry (Ville, 1990). As ships grew bigger, so did the size of shipyards and the scale of operations. The huge and weighty hull components forced yard-owners to expand their yards and use cranes to lift heavy materials. Many shipbuilders were unable or unwilling to make this change, and continued to build wooden ships. As a result, the centre of gravity in British shipbuilding moved north to the Clyde and the northeast of Britain (Harrison, 1990). New jobs with new skills and competencies emerged in shipyards, for example, metal and woodworkers, plumbers, electricians, millwrights and engineers. In the 1870s and 1880s many new machines were introduced in shipyards, for example, riveters and other kinds of machinery for working iron plates (Harley, 1973). The new machines often used new power sources. In the 1880s and 1890s, hydraulic power was introduced for heavy sustained work, while pneumatic power was used for more accurate and precision work (for example, drilling, planing and riveting). From the mid-1890s, electric power entered shipyards, powering power punches, boring machines, pumps, winches, saws, fans, cranes, drills, cutting machines and riveters (Pollard and Robertson, 1979). Another major transformation in shipbuilding was the introduction of science and engineering. While shipbuilding had always been a craft profession, it gradually turned into an applied science by the late nineteenth century. Professional naval architects and mechanical engineers entered the shipyard, substituting rule and calculation for instinct (ibid.). The naval architects, scientists and mechanical engineers were trained in formal institutions rather than on the job. A new education system was created, such as the Institute of Naval Architecture in 1857 and the subsequent creation of chairs of Naval Architecture at the universities of Liverpool, Newcastle and Glasgow (Ville, 1990).

A fourth element of the socio-technical regime that was substantially changed was user practices and market institutions. Users of steamships became large corporate and professional companies, offering new, regular liner services running to a regular timetable. A new kind of ownership (the joint-stock company) and new management practices were implemented. With regard to fleet management, the introduction of the submarine telegraph cable in the late 1860s and early 1870s had an enormous impact. It not only provided up-to-date information on prices in commodity and freight markets around the world, but also made possible much tighter control of ships. Information and instructions could be conveyed in hours rather than weeks. The telegraph enabled managerial activities to be concentrated in managing directors. Ships could be more easily directed away from markets with a surplus of shipping tonnage to those with a shortage. Fleet managers could assess alternative uses of steamships and send instructions to the captains to alter course. This led to a considerable increase in the efficiency of deployment of ships.

4.6 WIDER IMPACTS ON SOCIETY (1890–1914)

The transition to steamships not only affected the shipping regime, but also had wider effects on society. Steamships were indirectly involved in major social and economic transformations through their role in emigration. They transported great numbers of European emigrants in the late nineteenth and early twentieth centuries. These immigrants helped to fill the shortages in American labour markets, but also changed the social and political landscape in America. Because steamships were associated with emigration, they became symbols of hope and expectation, promising to transport people to a better future. Steamships became one of the icons of the industrial age. They were the largest moving objects ever made. They shrank the world, and by facilitating mass emigration they became the vehicles of social and political change.

Because steamships were icons of the industrial age, there was a prestige factor attached to having the biggest and fastest ships. As Germany grew into an industrial power, it wanted to have big ships to challenge Britain's supremacy. This led to a competition between both states in building the largest and fastest ship. Germany's emperor gave financial support to North-German Lloyd (Company) to build a series of large passenger ships, which could win the Blue Riband prize. The *Kaiser Wilhelm der Grosse*, launched in 1897, was the first of four ships of over 14,000 gross tonnage. It was 199.5 metres long, used two triple expansion steam engines and could travel at 22 knots. On its maiden voyage the ship captured the Blue Riband prize, establishing Germany as a major maritime power. Britain reacted and launched the 209-metres *Oceanic* in 1899. In response Germany launched the 17,000-ton *Deutschland* in 1900, seizing the Blue Riband with a run to New York of 5 days 15 hours and 5 minutes. With aid from the British government, the Cunard Line built the first superliners, *Mauretania* and *Lusitania*, both 32,000 tons. Launched in 1907, they were the largest and fastest ships of the world. With its four steam turbines the *Mauretania* (240 metres long) reached 25 knots, enough to hold the Blue Riband until 1929. This competition in shipbuilding raised tensions between both countries and was a foreboding of the First World War.

Steamships were involved in the great expansion of global trade in the second half of the nineteenth century. As they grew larger, faster, cheaper and more reliable, they lowered transport tariffs and boosted the emergence of a world market.

Steamships also affected agriculture. From the 1860s ever-larger quantities of cheap grain from America and the Black Sea appeared on European markets, transported by trains and steamships. Because of these cheap imports, food prices went down on European markets. This led to the agricultural crisis of the 1890s, which threatened the livelihood of many European farmers. Imports also changed agricultural practices in Europe. Cheap corn from America was used to feed cattle, contributing to the expansion of cattle farming. Guano from Latin America was used as a fertiliser in agriculture and stimulated productivity.

Eating patterns and health were also affected. Cheap grain lowered the price of bread. As efficient refrigerator ships were developed in the 1870s and 1880s, large quantities of frozen meat were imported into Europe from Australia and Latin America. Other perishables such as apples and butter followed (Jarvis, 1998). People could buy more food, which improved their health and standard of living. Luxury products such as meat, dairy products, sugar and fruit came within the reach of more people.

4.7 CONCLUSIONS

In this concluding section I shall follow the case-study protocol, described in Chapter 3 (Section 3.5). I shall test first the multi-level perspective and then the different patterns.

Testing the Multi-Level Perspective

In Chapter 3, two questions were formulated as proxies to test the usefulness of the multi-level perspective. How did niches emerge in the context of the existing regime? To which ongoing processes at the regime and landscape levels did the niche link up? Both questions will be addressed below.

An important claim of the multi-level perspective is that new radical technologies emerge in specific niches in the context of the existing regime and landscape. This claim was substantiated in the case study. The first technological niche for steamboats was situated on inland waterways. In the context of the canal boom, since the 1780s experiments were conducted with steamboats on rivers and canals. The first market niche was created on the Hudson River (1807). On American inland waterways the steamboat was widely used in the 1810s and 1820s, because it linked up with the landscape process of westward settlements. Another niche was in ports, where the steamboat was used as a steam tug to help manoeuvre large sailing ships. In the 1820s there were experiments to add steam engines to existing oceanic sailing ships. They were used incidentally as auxiliary devices (add-ons), when there was no wind. They thus linked up with a problem in the sailing ship regime. In 1838 the British government created a new niche, mail transport. The government decided to subsidise steamships to transport mail within the Empire. This was a response to another problem in the regime: limited coordination in trade and politics. Because steamers transported mail more quickly, coordination was improved.

The other claim of the multi-level perspective was that the regime should not just be seen as a barrier for radical novelties. Established regimes may be stable, but are not inert. In socio-technical regimes and landscapes there are ongoing processes which may provide windows of opportunity to which niche innovations can link up.

The case study did indeed show that steamships linked up with several ongoing processes in the sailing ship regime. First, professional ship owners emerged as a new social group in shipping in the late eighteenth and early nineteenth centuries. Professional ship owners were the first group to adopt steamers and experiment with line services. Second, new institutions were created in trade in the early nineteenth century, for example, new trading actors, commercial journals, trade journals and commercial law. This institutional framework provided the context for the expansion of global trade in the 1830s and 1840s. Steamships linked up with this expansion and strengthened it. Third, steamships linked up with the problem of limited communication and coordination in trade. Stimulated by mail subsidies, steamships greatly reduced this problem in the 1840s by speeding up mail transport in the Empire. Fourth, steamships linked up with the problems of limited regularity and low predictability. Sailing packets, with fixed departure times, were a first step towards greater regularity. In the 1840s, steamships linked up with this trend and greatly reinforced it. With line services they introduced a new functionality in shipping, fixed departure *and* arrival times. Fifth, steamships linked up with market dynamics in the regime. Markets for passenger transport had already begun to grow in the 1820s and 1830s (Australia, America, Canada). In the late 1840s this growth speeded up. In particular, emigration to America provided space for the take-off of steamships. Sixth, the Suez Canal (1869) started as an external political project for prestige and power, but it greatly influenced shipping, opening the Indian and Chinese markets for freight shipping with steamships. Seventh, the shift from wood to iron linked up with the guiding principle of building longer ships and the problem of longitudinal strength. This problem created an opportunity for the introduction of iron. Composite sailing clippers were built in the 1850s, consisting of an iron frame and wooden planking. In the late 1850s and early 1860s, ships with all-iron hulls and steel masts became more common. Steamships could coast along with this general shift towards iron hulls and benefit from it, because iron could better stand the vibration from the screw propellers.

In the transition, external landscape developments also played a role, and steamships linked up with several of them. First, steamboats linked up with the colonisation of America and the westward settlements in the early nineteenth century. Second, industrialisation and liberalisation were two macro trends which stimulated the growth of trade and shipping markets. This created space and money for experimentation with new technologies (for example, iron hull, screw, compound engine). Steamships could coast along with these growing markets. Third, European emigration, stimulated by the Irish potato famine, the European revolutions of 1848 and the Californian gold rush of 1848, was a landscape development, which boosted the market for passenger transport. This was the first commercial market in which steamships took off.

In sum, I conclude that the multi-level perspective has a good match with the case study, and therefore is robust.

Co-Evolution of Multiple Technologies

The system innovation in shipping took place because of interactions between multiple technologies. One pattern, already mentioned in Chapter 3, was the importance of incremental innovations alongside radical innovations. In the 1860s a new steamship regime emerged because three technical trajectories merged: screw propulsion, iron hull and compound engine. These trajectories were accompanied by a range of incremental innovations in the ship. The use of iron hulls, for instance, required adaptations in the compass, and anti-fouling paints to prevent marine growth. Improvements in steam engines depended on improvements in boiler design (for example, tubular boilers), better lubricants, steel (to build stronger boilers), compound engines and surface condensers. And the use of screw propellers required iron hulls to solve the vibration problem, gearing mechanisms to achieve a higher number of revolutions and stern gland fittings.

A second pattern was complementarity between technologies. Some technologies could only break through after a complementary technology was developed: surface condensers, for instance, were a complementary technology for the diffusion of compound steam engines on oceanic steamers; iron hulls were a complementary technology for the screw propeller, which had a higher speed of rotation than paddle wheels.

Third, (unexpected) cross-technical linkages positively influenced each other (not strictly in a technical sense). In the 1870s the functioning of the shipping regime was greatly improved with the international telegraph. Until then, mail was the only means of long-distance communication. Because the telegraph reduced long-distance communication to minutes, it greatly impacted coordination and management in trade and shipping. The telegraph thus stimulated professional and centralised fleet management in shipping.

Fourth, symbiotic relationships developed between old and new technologies. Steamships and sailing ships not only competed, but also had a symbiotic relationship in a technical and functional sense. In a technical sense, there was a pattern of add-on and hybridisation. The first steamships were actually sailing ships with additional steam engines. Steamships in the 1840s were hybrid forms with both sail and steam propulsion. In a functional sense, early steamers improved the functioning of the sailing ship regime. Steam tugs helped large sailing ships manoeuvre in ports and mail steamers improved coordination in trade. Even the breakthrough of steamers had symbiotic aspects, because it created a new market niche for sailing ships: worldwide coal transport.

Fifth is the importance of material innovations. Anti-fouling coatings were developed for iron hulls. Iron and steel were important for shipbuilding because they allowed stronger and much larger ships. Steel was also important to build stronger boilers, which could withstand higher pressures.

Diffusion as a Trajectory of Niche Accumulation

Steamships diffused because they were used in subsequent application domains. Steamboats were first used on inland waterways and for minor functions such as towboats, then as subsidised mail steamers, then for passenger transportation, and eventually also in freight shipping. Figure 4.11 gives an overview of the process of niche accumulation.

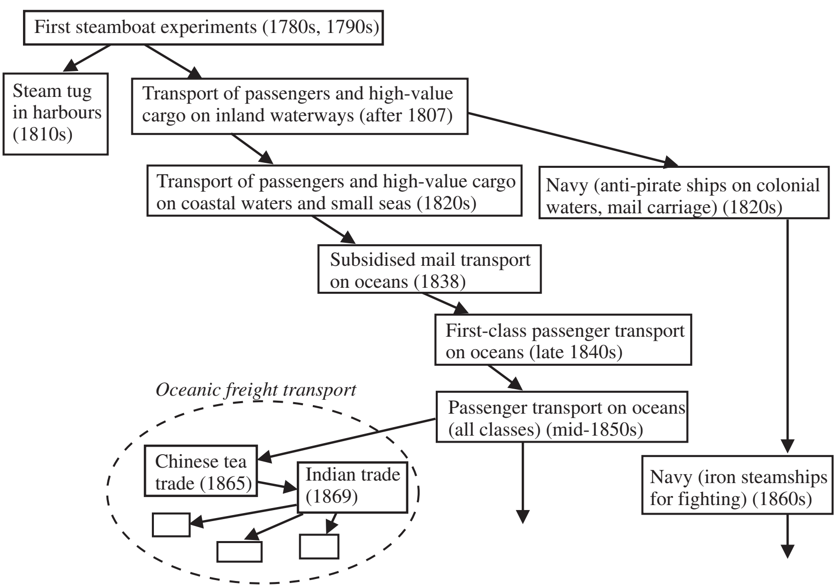


Figure 4.11 Trajectory of niche accumulation for steamships

Phases in the Innovation Journey of Novelty

In Chapter 3, I distinguished four phases in system innovations. In the steamship transition these four phases did indeed occur, and therefore, I conclude that they are corroborated.

First phase: emergence of novelty in the context of existing regime and landscape

The new technology was fitted within the existing regime to solve particular problems. The first steamships did not compete with sailing ships, but were used as steam tugs to manoeuvre large ships into ports. After 1838 steamships were subsidised and used to transport mail, thus improving long-distance communication. Also in a technical sense the new technologies were fitted in the existing regime. The first oceanic steamships were little more than sailing ships with additional steam engines and paddle wheels.

Second phase: technical specialisation in market niches; exploration of new functionalities

In the late 1830s, steamships were used on oceans for subsidised mail transport. This new application domain led to particular problems (high coal consumption, unstable paddle wheels in high seas, buckling of the wooden hull under heavy steam engines). Dedicated engineers and shipbuilders directed their attention to these problems, and developed new technologies, for example, screw propulsion, iron hull and compound engine. This led to the emergence of a dedicated community of specialised steamship builders. Experiments with new technologies (for example, iron hulls) were improvised out of existing shipbuilding rules. New technical forms emerged, in particular hybrid forms of sailing ships with steam engines. In the late 1840s, the new steamships were used in the expanding market niche of passenger transportation. This led to the articulation of a new functionality: line services. The new functionality required learning on the part of users, for example, new management practices such as fleet management and financial planning.

Third phase: breakthrough, wide diffusion and competition with the established regime

In the 1850s the use of steamships in passenger transportation expanded. In this domain, they quickly replaced sailing ships. In freight shipping the diffusion was slower. British steamers entered trades with northern Europe and the Baltic countries in the 1850s, and the Mediterranean fruit trade around 1865. With the introduction of the compound engine in the mid-1860s, steamships were used in the Chinese tea trade. At the end of the 1860s, freight steamers were used in the North Atlantic grain trade.

Disciplinary perspectives give different explanations of this process. *Economic perspectives* emphasise that the diffusion process was driven by price/performance improvements. Improvements in steam engines led to major reductions in coal consumption: from 5 lb per horsepower per hour in 1855 to 1.5 lb in 1890. Other aspects also played a role in reducing operating costs of steamers, for example, scale economies, reduction in crew size and decreasing prices for iron hulls. The *multi-level perspective* emphasises the importance of changes in the external context, for example, European emigration, industrialisation and the opening of the Suez Canal. *Socio-technical perspectives* emphasise that diffusion took place because an increasing number of heterogeneous elements were linked together. A new technical steamship regime emerged in the late 1860s as three technical trajectories linked up: screw propulsion, iron hulls and compound engines. Market dynamics were also important, as were political subsidies, adaptations in ports, the construction of a worldwide coal infrastructure, new sources of capital (joint-stock company) and so on. *Sociological perspectives* highlight particular social mechanisms, for example, the bandwagon effect involved in the steamship mania after the opening of the Suez Canal, the subsequent backlash, the sailing ship effect and social struggles about the introduction of bulk cargo-handling machinery (because of fear of unemployment). Sociological perspectives also highlight all kinds of rules and perceptions, and how they change over time (for example, perceptions about iron ships, design rules for building iron ships, problem agendas, search heuristics, insurance rules, subsidies). None of these disciplinary perspectives offers the 'best' explanation; they each highlight different relevant aspects.

Fourth phase: gradual replacement of established regime, transformations and wider impacts

As the performance of steamships was improved, their operating costs went down. Sailing ships were increasingly replaced in freight markets. The wide diffusion and replacement was accompanied by changes in shipyards, ports, quays, cargo-handling equipment, coal infrastructure and so on. Sailing ships defended themselves with technical improvements (the 'sailing ship effect') and transfer to other market niches (freight transport in bulk markets).

Steamships had wider impacts on other societal domains, for example, globalisation and worldwide trade, agriculture, eating patterns and health, and emigration. These impacts were usually produced because steamships linked up and accelerated ongoing processes.

Table 4.10 Fit–stretch pattern in the co-evolution of form and function in the steamship transition

Technical form / Use Environment	Fit		Stretch
Fit	Early steamships are sailing ships with additional steam engines, used as tugs in ports		
		Steamships are hybrid forms with both sails and steam engines plus paddle-wheels. Used for mail transport and passenger transport in the 1840s. New functionality was articulated: line services	Ships became steamships with additional sails. New technical forms are created in the 1860s, through linking of screws, iron hulls and compound engines. Steamship break through in freight transport in 1869 (Suez Canal)
Stretch			Diffusion in freight in 1870s and 1880s. Accompanied by wider adjustments in ports, shipyards, loading equipment, infrastructure. Steamships have wider impacts

Fit–Stretch Pattern in the Co-evolution of Form and Function

In a fit–stretch table, I have summarised the co-evolution of form and function during the steamship transition (see Table 4.10).

NOTES

1. Howells (2002) has recently cast doubt on the prevalence of the sailing ship effect. He argues that sailing ships and steamships were employed in such different market segments that they did not really compete, and that improvements in sailing ships may have been the outcome of competition between sailing ship firms.
2. Van Driel and Schot (2001) describe how the introduction of grain elevators in Rotterdam harbour led to a struggle among labourers, merchants, ship owners and elevator producers.

5. The transition from horse-drawn carriages to automobiles in American urban passenger transportation (1860–1930)

5.1 INTRODUCTION

This chapter analyses the transition from horse-drawn carriages to automobiles in America from 1860 to 1930. This transition involved changes in technology, not only horse-drawn carriages and automobiles, but also bicycles and electric trams. It also involved changes in infrastructure, for example, rail infrastructure for horse and electric trams, and also road infrastructures with new kinds of surfaces (asphalt, concrete). New types of roads were created with limited access for a particular kind of vehicle (highways). In addition, the responsibility for the administration of roads shifted from local residents to public authorities (highway engineers, urban planners). In fact, the symbolic perception of the role of the street changed from social meeting place to transport artery. New traffic rules were formulated to create more order on the street. Pedestrians were moved to pavements and slower traffic to the side of the street. The street was increasingly organised around the automobile. While transport in 1860 was mainly public transport along fixed routes (for example, omnibus, horse tram), the automobile introduced new functionalities in the transport system, namely individual transport (driving yourself) and flexible transport (chose your own route and time of travel). These new functionalities were widely explored and enjoyed in new application domains, for example, racing and touring in the countryside. The car also allowed urban middle classes to move out of the city and enjoy suburban lifestyles. In sum, the technological transition to automobiles was part of a wider system innovation. This chapter analyses how the transition and system innovation came about.

The analysis has a focus on *urban* transportation, because new transport technologies (bicycles, electric trams, and automobiles) were first developed and used in cities and then spread wider. The story also has a focus

Table 5.1 Annual world motor vehicle production, 1900–1924

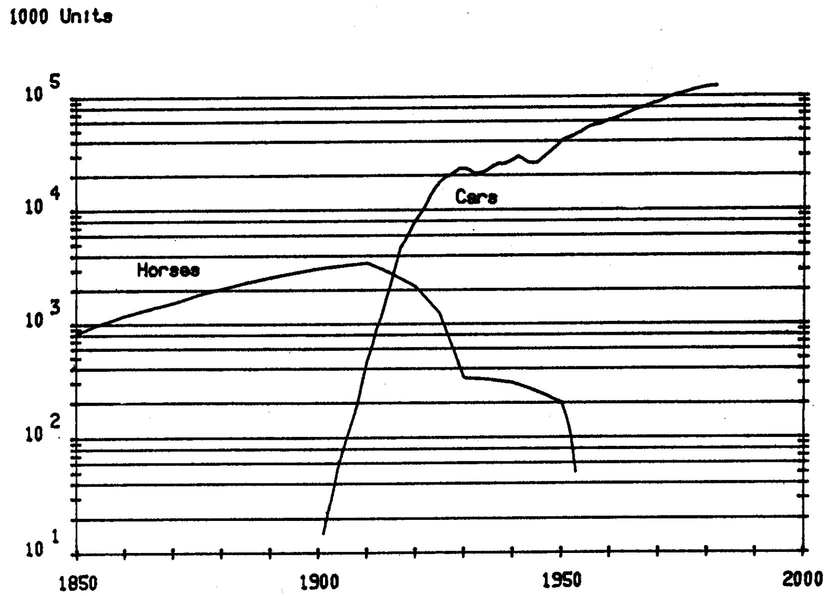
	1900	1907	1913	1924
United States	4,000	45,000	485,000	3,504,000
France	4,800	25,000	45,000	145,000
United Kingdom	175	12,000	34,000	147,000
Germany	800	4,000	23,000	49,000
Canada	–	–	18,000	135,000
Italy	–	2,500	8,000	50,000

Source: Ville (1990: 180).

on *passenger* transportation. I shall pay less attention to the transition in freight transport, which figures mainly in the story as an application domain, where horses continued to be used long after they had been driven out of passenger transportation. The analysis focuses on America, because this country was the leader in the shift to automobiles (see Table 5.1). Only before 1900, did other countries, such as France and Germany, take a lead in automobile developments.

The transition from horse-drawn carriages to automobiles is often portrayed as a substitution process in which horses were replaced by cars (see Figure 5.1). The emergence of new technologies is assumed to be at the expense of old technologies. Nakićenović (1986: 316), for instance, writes that: 'In the US the first horseless carriage posed an alternative to the horse-drawn buggies and wagons. Especially as a commercial vehicle, the motor car offered many advantages', for example, increased radius of local transportation. And Grübler (1998: 64) writes: 'the car industry grew initially by replacing horses'. Old and new technologies are thus assumed to compete right from the start. But the analysis in this chapter will show that the dynamics were more complex. In fact, the substitution approach is a distortion of historical reality. Urban horses were first replaced by the electric tram, and only later by the automobile. Although the electric tram was replaced by the automobile and the bus in the 1930s and 1940s, that is no reason to dismiss its importance.

Furthermore, a simple substitution approach is insufficient, because it looks mainly at technical aspects and markets. This kind of analysis neglects other aspects, which were involved in the system innovation, for example, infrastructures, traffic rules, symbolic meaning and new functionalities. To analyse the interplay between heterogeneous aspects, I shall use the conceptual perspective described in Chapters 1 and 3. First, I shall give an indication of the socio-technical system in transportation, to get an impression of



Source: Nakićenović (1986: 320); courtesy of Elsevier Science.

Figure 5.1 Population of road horses and number of cars in the USA

the relevant elements. Figure 5.2 presents a sketch of the system, filled in for automobiles. The analysis in this chapter will show how these heterogeneous elements interacted with one another, how changes in one element triggered changes in another, and how developments gradually linked up.

I also present a figure of the main social groups involved in urban transport (Figure 5.3). These social groups maintained (and changed) the elements of the socio-technical system. They acted in the context of a set of rules, the socio-technical regime, which involved design rules, problem agendas, search heuristics, exemplars, tax rules, user practices and so on.

These indications of system, social groups and regime show that the transition was complex, involving many elements, actors and rules. The question is how interactions between them resulted in a system innovation. To structure the analysis I shall use the multi-level perspective, described in Chapter 3. This is a very useful perspective to analyse this case study, because America in the late nineteenth century was a society in flux. The empirical description explicitly analyses the transition in its wider societal context, because immigration, urbanisation and suburbanisation were macro changes on the social dimension. The emergence of entertainment,

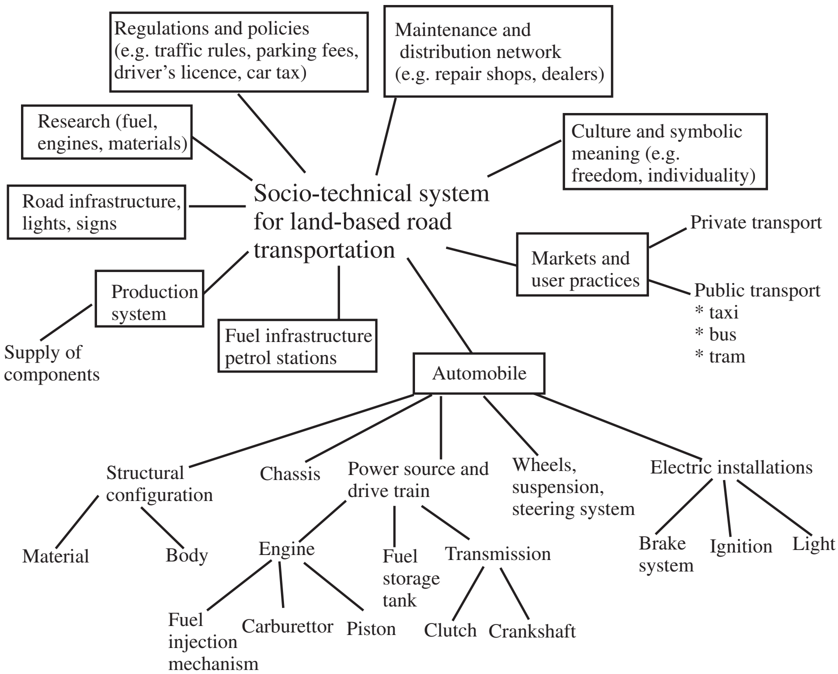


Figure 5.2 Socio-technical system in personal land-based transportation

consumption and excitement as new values were macro changes on the cultural dimension. Reform movements and the expansion of bureaucracies constituted macro changes on the political dimension. These changes influenced the transition in many ways, and the influences can be analysed using the multi-level perspective. The hypothesis is that the interaction among niche, regime and landscape levels explains the timing and course of the system innovation. This hypothesis will be tested in the concluding section of the chapter.

Another aim of the chapter is to test the patterns described in Chapter 3: co-evolution between multiple technologies, diffusion as a trajectory of niche accumulation, phases, and fit–stretch pattern in the co-evolution of form and function. These patterns will be explicitly addressed in the concluding section.

The period for analysis stretches from 1860 to 1930. This means that I do not start with the emergence of automobiles in the 1880s. Rather than recounting the grand story of the automobile in isolation, the aim is to show that automobiles were the last stage in a much wider transformation process that had already begun in the 1860s and 1870s when, because of internal

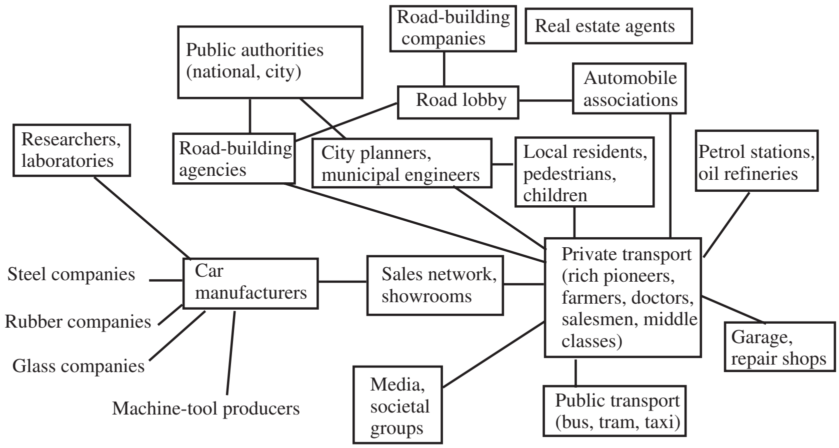
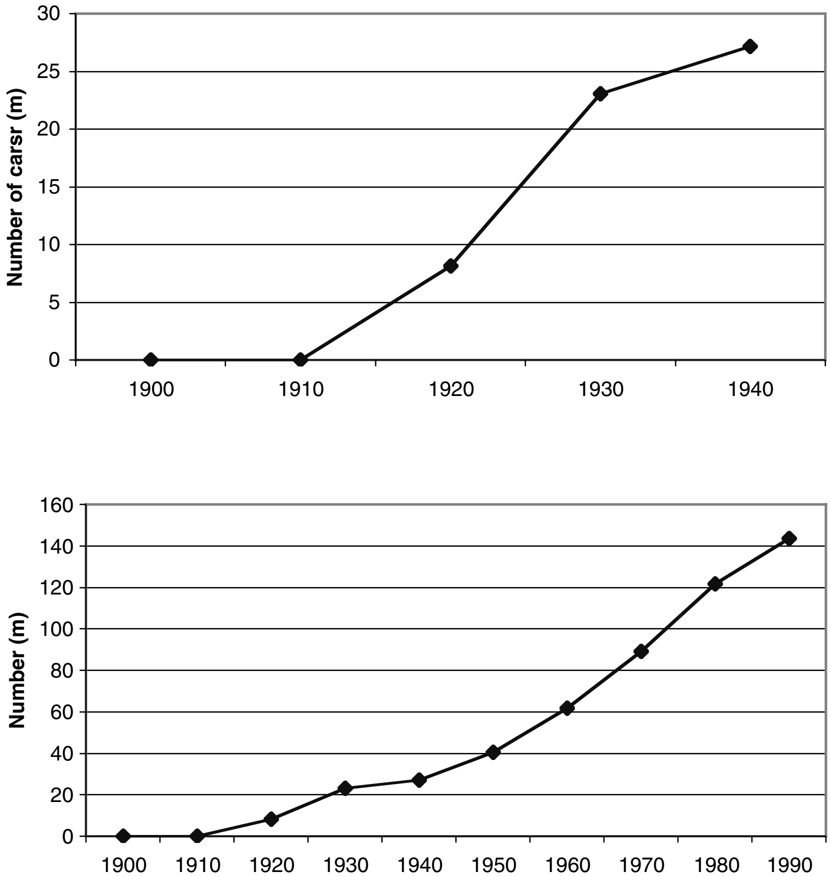


Figure 5.3 Social groups in personal land-based transportation

pressures, the existing horse-based urban transportation regime opened up. In response to this opening up, new transport technologies emerged in the 1880s and 1890s, for example, bicycles and electric trams. Electric trams in particular played an important role in terms of overall mobility, while automobiles emerged in some small niches and played a minor role. Automobile technology gradually developed over time, for example, the T-Ford was the dominant design in 1908, and began to overtake the electric tram in the 1920s and 1930s. The main point is that the transition did not start with automobiles, which should rather be seen as the final ingredient in a much longer process. The empirical story ends in the 1930s, not because the transition was entirely completed, but because the general direction of further developments was relatively clear. The electric tram was in strong decline, while the automobile was widely seen as the technology of the future. It is also appropriate to stop the analysis in the 1930s, because in this decade the first diffusion curve levelled off. The diffusion of the car in America can be divided into two S-curves, the first from 1900 to the 1930s and the second from the Second World War to the present (see Figure 5.4). The first S-curve represents the shift from technical novelty to culturally accepted transport mode. By the 1930s, it was widely thought that the automobile represented the way to the future.

The empirical description is organised along the four phases in system innovations, describing different periods. The first period (1860–85) describes the 'heating up' in the urban transportation regime (Section 5.2). While the horse tram was a major innovation in the horse-based transportation



Source: Based on data from Mowery and Rosenberg (1998: 50).

Figure 5.4 Automobile registrations in America, 1900–1940 and 1900–1990 (m)

regime, some new technologies were pioneered and used in particular niches, for example, bicycles, electric vehicles and steam automobiles. The second period (1885–1903) is characterised by fluidity and transformation in the transportation regime (Section 5.3). Pressure in the horse-based regime was increasing because of internal and external problems and changes in cultural values. This pressure stimulated the diffusion of the bicycle and electric tram, which played a role in the articulation of new user preferences, new mobility practices, new social lobby groups, social struggles, new infrastructures and new cultural perceptions. The 1890s were a crucial decade, because many new elements were introduced in the socio-technical regime. These elements would later coalesce around the gasoline vehicle in the early twentieth century. The third period (1903–14) is characterised by the expansion of the electric tram and the stabilisation of the gasoline car in the T-Ford (Section 5.4). The fourth period (1914–30) is characterised by the rising popularity of the gasoline automobile and the decline of the electric tram (Section 5.5). A new socio-technical regime was created around the automobile, which had wider impacts on society. The chapter ends with conclusions (Section 5.6). For each section the following dynamics are described:

- Relevant external landscape developments.
- Main developments in the socio-technical regime, for example, technological developments, market dynamics, policy actions, problems and tensions. How were these regime processes influenced by landscape developments?
- Emergence and development of novelties in niches, and the actors involved, the problems they encountered, and their attempts to overcome them. How were niche dynamics related to ongoing processes at the regime and landscape levels?

5.2 EXPANSION OF HORSE-BASED TRANSPORTATION (1860–1885)

Landscape Developments

At the landscape level, industrialisation and urbanisation were two important macro developments. The emergence of steam power enabled a relocation of factories to cities, where railroads provided cheap coal. This relocation stimulated urbanisation. By 1890, more than 30 per cent of all Americans were city dwellers (Faragher et al., 1997: 596). The growth of cities led to longer travel patterns, which were hard to undertake by foot.

The growth of cities was accompanied by problems, particularly for poor working-class people who lived in crowded and filthy slums. The swelling urban population produced human excrement in unprecedented quantities and much of this found its way onto public streets. Health professionals declared that the decaying organic filth filled the air with poisonous vapours, 'miasmas', which were seen as the cause of epidemic diseases (for example, cholera, yellow fever, tuberculosis). Health and hygiene became important cultural issues, supported by a hygiene movement. The concern about hygiene would have a negative influence on horses, which were seen to contribute to the filthy streets. But cycling and car driving would link up with this concern positively, because they were seen as healthy.

Light and fresh air were seen as solutions to filthy living conditions, improving public health. Hence, parks were built to function as the lungs of the city. Rich people also drove around the parks in their private horse-drawn carriages, a popular pastime at evenings or weekends (McShane, 1994: 30–31). This was the beginning of a new mobility practice: 'driving for fun'. The parks later became arenas for the demonstration of new transport modes.

The general public came to see cities as filthy, unhealthy and dangerous. Many middle-class American families began to see suburban living as desirable, a haven from the tumultuous city (ibid.: 23). Thus, suburban living emerged as a cultural ideal in the United States before it became feasible with several transport innovations.

Another macro change was the emergence of a middle class, consisting of salaried employees, managers, technicians, clerks and engineers. The new middle class had more money and more leisure time, to be enjoyed in the form of entertainment (Faragher et al., 1997: 605). A new popular culture emerged in America, valuing such things as entertainment, excitement, fun and sporting activities. Early cycling and car driving would later link up with these new values.

Innovations in the Urban Horse-based Transport Regime

Urban transportation was a dynamic and innovative regime in the second half of the nineteenth century. As cities grew in size, so did travel distances. In response, several transport innovations were developed, mainly for the middle and upper classes. One innovation was the omnibus, a 'rectangular box on wheels', carrying between 15 and 25 people. It was drawn by two horses and was operated on fixed routes at scheduled intervals, picking up anyone who flagged it down and dropping off passengers where they chose. New York had 70 omnibuses in 1830, 350 in 1849 and a peak of 683 in 1854 (ibid.: 8). Travel fares were too high for the mass of the city population, but vast numbers of better-off people made increasing use of these horse-drawn vehicles (Barker and Gerhold, 1993: 67). For wealthy people there were also more private forms of (semi-)public transport: four-wheel hackney vehicles and two-wheel cabriolets (or cabs).

An important innovation was the horse tram, that is, horse-drawn carriages on railroads. This way a horse could pull a vehicle at 30 per cent higher speed and haul 100 per cent more passengers (McShane, 1994: 14). The horse tram helped to expand the city limits and enabled lower travel fares. It replaced the omnibus on busy routes in the 1850s and 1860s and became the first urban mass transportation mode. The horse tram helped to create the first wave of middle-class suburbanisation. In 1880 there were about 19,000 horse-tram wagons in America, drawn by about 100,000 horses and operating on an infrastructure of 4800 kilometres (Mom, 1997: 29). Horses pulling a tram were replaced every 2–3 hours. Large horse-tram companies operated fleets of thousands of horses, which were housed in large stables. This entailed risks in the event of disease and fire, so a wave of stable regulations were instituted. Thousands of stable boys, blacksmiths and carriage manufacturers were involved in feeding and stabling the horses, and maintaining the carriages. Feeding and stabling accounted for more than 50 per cent of the tramway companies' costs (Chant, 1999: 130). City authorities established sanitation departments to clean the streets of horse excrement. In sum, an entire regime was created around horse-based transportation, involving many social groups, and this regime continued to expand until the end of the nineteenth century.

Horse trams, carriages, wagons and pedestrians all mixed in the streets. Although streets were used for transportation, they also functioned as social meeting places, for play, social talk and street vending. They were unpaved dirt roads or covered with rough cobblestones. Horse-drawn vehicles found cobblestones to be problematic, because the cobbles damaged the wheels. Hence, there was pressure to make the road surfaces more even. However, this did not always have the desired effect, because many streets were administered by local residents with responsibilities for deciding about maintenance and improvement. Local residents had no interest in improving street surfaces, because for them, traffic was more of a hindrance than a benefit. Many residents felt that smooth roads were for the rich, not for average citizens (McShane, 1994: 64). Hence, they often withstood pressure to improve the road surfaces. Some local residents actively chose to neglect the roads. But on busy avenues and on park drives a first wave of street improvements began after the Civil War (1861–65). These road improvements used macadam, that is alternating layers of stones, large ones at the bottom, smaller ones on top, and covered with gravel; when the roads were used, the layers of stones were compressed and became very robust. Macadam roads had a relatively smooth surface and low rolling resistance. They were developed for horses and light vehicles, for which they provided easy traction. The street improvements encountered setbacks in the early 1870s, because of financial scandals, overspending and corruption, and fear of mismanagement slowed expenditures on public improvements until the 1890s (ibid.: 69).

New Transport Technologies in Niches

New transport technologies emerged in particular niches. The earliest bicycles were developed in the 1830s as toys for the upper classes. At the end of the 1860s a new application domain was opened up: bicycle racing on racetracks. The users were athletic young men, wanting to prove their skill, courage and strength. These user preferences gave rise to the penny farthing or 'Ordinary', a bicycle with a very large front wheel. This enabled high speed, but was very unstable. However, this instability was not seen as a problem, because the bicycle functioned as an 'adventure-machine' (Mom, 1997). In fact, this characteristic was appreciated, because it heightened the audacious, daring and skilful image of the bicycle (Pinch and Bijker, 1987).

Another radical novelty was the electric automobile. The first experiments go back to 1842, when Robert Davidson constructed a four-wheeled carriage that was nearly 5 metres long, weighing about 5 tonnes (Bowers, 1995: 378). These early electric vehicles were handicapped by weak electric motors and heavy batteries.

There were also attempts to use steam engines for propelling vehicles. Steam buses were commercially used in London in the early 1830s. But steam engines were heavy and cumbersome, and prone to boiler explosions. In the 1860s, steel made possible new boiler designs and lighter steam automobiles. In America, this resulted in a new design trajectory: smaller and lighter steam vehicles. Sylvester Roper, for instance, built a light (650 pound) steamer in 1863, capable of doing 20 miles per hour. In the 1860s and 1870s, steam automobiles were used as fascinating novelties at racetracks and in circus parades (McShane, 1994: 92). Although further technical improvements were made in the 1870s and 1880s, steam automobiles ('steamers') did not catch on, because of public resistance and regulation. Regulators reflected public opinion when they banned steamers because of their speed, smoke and potential for explosion (ibid.: 97). Because people had not (yet) become accustomed to high speeds, they opposed steamers in their streets.

Other innovative efforts were directed at finding mechanical ways to propel trams. Steam trams achieved moderate success in cities, because of public resistance against smoke and potential explosions. Steam was also used in cable cars, which were pulled by (underground) cables powered by a central steam engine. Electricity also seemed a promising candidate. Between 1879 and 1888 many electric trams were displayed at expositions and exhibitions.

5.3 HEATING UP AND FLUIDITY IN TRANSPORTATION (1885–1903)

Landscape Developments

At the end of the nineteenth century, America was a society 'in flux'. Several landscape developments from the earlier period continued, such as the growth of industrial cities, the emergence of a new middle class, the rise of a new popular culture (with values such as excitement, entertainment, adventure and speed) concern about health and the hygiene movement. An important new landscape development was immigration. In the 1880s and 1890s, about 600,000 people entered America per year, mostly settling in cities (Faragher et al., 1997: 596). The majority of urban populations lived in crowded tenements. The quality of life in cities was poor: overcrowded conditions and inadequate sanitary facilities bred contagious diseases, and public health remained a major concern. The city centre became more congested and noisy.

Another macro development was the rise of political reform movements. The period between the 1890s and the First World War is often characterised as the 'progressive era'. Progressivism was a varied collection of reform communities, aiming to end political corruption and address poverty and urban problems. Many urban social reformers believed that suburbanisation would eliminate social problems associated with slums. Because they saw electric trams (and later automobiles) as a means to this end, they encouraged their diffusion (McShane, 1994: 77). In the 1890s, the cultural ideal of a suburban home became reality for the middle classes.

A related macro development was the expansion of public administration at all levels (municipal, county, state and federal). Public authorities took greater responsibility for regulating society. The new jobs were filled mainly by the new middle classes, who brought their own values and perceptions with them. These greater public responsibilities sometimes led to struggles with other actors. For instance, with regard to the choice of road surfaces, public authorities clashed with local residents, and litigation and the courts became a common way to settle disputes. To demonstrate their new power, cities encouraged the creation of impressive structures. This was called the 'City Beautiful' movement (Faragher et al., 1997: 599). Grand concrete boulevards were constructed at high public cost and individual parks were connected with smooth paved 'parkways'. The so-called 'Parkway' movement climaxed in the 1890s. Riding with carriages on parkways gave rich people the exciting experience of high speed (McShane, 1994: 40). Parks also created a new arena for the demonstration of new carriages and vehicles.

The creation of parks and boulevards was part of a broader trend in the late nineteenth century to segregate and reform public space (Baldwin, 1999), which was divided into various restricted public spaces. Streets were cleaned up, public advertising suppressed, selling on the street was forced off the thoroughfares and into the slums, and recreational activities were pushed into parks, playgrounds and side streets. These reform efforts turned public streets into a battleground for social and political power. Local residents resisted the middle-class reform campaigns, but the middle classes gained the upper hand as they acquired powerful positions in city governments and judicial bodies.

In the late nineteenth century, electricity was experienced as the symbol of a new age (just as computers were in the late twentieth century). The cultural optimism about electricity stimulated investors to put their money into electric technologies, for example electric trams, electric light and electric automobiles.

In rural areas there was much discontent about the railways. Farmers blamed railroad companies for charging exorbitant fees for transport services. Railroads commonly charged farmers far more to transport their crops over short distances than over long hauls. Hence, the railroad industry became the focus of rural protest. This provided fertile soil for the automobile in later years. Rural areas also suffered from urbanisation: many young people left the countryside to work in cities, thus eroding rural institutions such as schools, shops and medical facilities (Ling, 1990).

The Horse-based Transport Regime: Expansion and Problems

Between 1870 and 1900 the use of horses for transportation grew exponentially, in the form of horse trams, freight wagons and horse-drawn taxis. But wider landscape developments increased the pressure on the transport regime, leading to several problems. The first was the high cost for tram companies (especially for feeding and stabling horses). This provided an incentive to look for alternative ways of propelling trams, for example, with steam or electricity.

The second problem was congestion. Horses and wagons occupied much street space. It was increasingly difficult for streets to process the increasing number of wagons and carts. But the problem was also due to the fact that different transport modes operated at different speeds: a few slow vehicles could easily create major traffic jams. Alternatively, congestion was caused by overworked horses, which sometimes dropped dead in the middle of the traffic. It took the sanitation department hours to remove a dead horse.

Third, there was the problem of safety: horses kicked and bit and people were run over by carriages, most of the victims being pedestrians. In New York the number of people killed by horse wagons and streetcars rose from about 4 per 100,000 inhabitants per year in 1885 to 11 in 1900 (McShane, 1994: 49). This indicates a rapid rise in the number of accidents in the 1890s.

The fourth problem was pollution. Horses produced huge amounts of excrement, over 150,000 tons annually in New York (ibid.: 52). Frequent urination added to the mess, resulting in soiled city streets, which stank terribly in the summer. This was thought to result in dangerous miasmas, which spread disease. Hence, horses acquired a negative public image. In response to this health problem, municipalities expanded their city cleaning departments and employed more street sweepers.

These problems of the horse-based transportation regime created windows of opportunity for new transport technologies, which were developed in niches. The electric tram broke out of its technological niche and rapidly conquered a market share, replacing the horse tram. The bicycle found a market niche with the middle classes who used it for recreation and touring in the countryside. And urban authorities used their enhanced authority to push through road improvements at the expense of local residents. In sum, the 1890s were a hectic period in which many new transport configurations competed for attention, investors and consumer preferences. As a result of this increasing variety, the urban transport regime opened up and became more fluid. Implementation and use of the bicycle and tram gave rise to transformation processes, which led to changes in mobility practices, urban layouts and perceptions of the function of streets. In this context, there was also a new experimental transport option: automobiles. However, this new option was still in a technical pioneering phase and did not play a significant role in terms of overall mobility. These niche-development processes and their wider transformation effects are described in the next subsection.

Trams, Bicycles and Automobiles in Market and Technological Niches

In the early 1880s electric trams were technical novelties, demonstrated at expositions and exhibitions. But in 1888 the first market niche for electric trams was created in hilly Richmond. The subsequent diffusion of the electric tram in America was very rapid. In 1890, 16 per cent of American street railways were electrified, about 70 per cent were horse or mule powered, and 14 per cent consisted of cable cars or steam railways. By 1902, 97 per cent of American street railways were electric (Hilton, 1969: 126).

There were several reasons for this very rapid rise of electric trams. First, the electric tram had operational advantages: it was about twice as fast as the horse tram (12 mph versus 6 mph), and it eliminated the problem of tons of horse excrement. Also, electric trams were cheaper, reducing the cost per car-mile from 8 to 11 cents for horsepower to 1.5 cents for electric power (Hunter and Bryant, 1991: 209). A problem, however, was that high upfront costs had to be paid to construct new infrastructures (for example, electricity wires, stronger rails). In addition, the purchase of electric trams was a substantial investment. However, the price of electric trams decreased over time, from $4500 in 1889 to $750 in 1895 (Mom, 1997: 41). A second reason was that electric trams benefited from the cultural enthusiasm for electricity. It was not difficult for tram companies to find willing investors. A third reason was that powerful social groups supported the electric tram: horse-tram companies wanted to discard their expensive horses; real estate promoters invested in tramlines because it increased the value of their land (Nye, 1998: 166); electric light companies stimulated electric trams, because it provided them with a daytime market; and progressive policy makers saw low-cost electric trams as a means to stimulate suburbanisation, thus eliminating the slums (McShane, 1994: 29).

The trolley constituted a transportation revolution in the sense that it created the first urban system of mechanised mass transit. The fare of tram trips was relatively low, five cents, with free transfers (Nye, 1990: 96). This made mechanised transport possible for many people. In 1902, when there were only a few thousand automobiles in the country, Americans took 4.8 billion trolley trips and made one billion transfers (ibid.); because their costs were low, tram companies made high profits in the 1890s (p. 90). As the electric tram diffused, it had wider transformational effects. First, it resulted in the elimination of horse trams. Second, the electric tram greatly contributed to suburbanisation. The rich and the middle classes moved to suburban areas, while the working classes remained in the city centres. Because the electric tram was faster and travelled further than the horse tram, vast amounts of land were opened up which could be developed for housing by real-estate speculators (see Table 5.2). Third, the electric tram helped to transform the city centre, for example, by allowing the concentration of business and entertainment. Cultural activities such as theatres, museums and the cinema were increasingly situated in the city centre, as was shopping in department stores (Rae, 1968b: 302). Fourth, electric trams stimulated tourism for the lower classes, for example, day trips, picnics or walks in the country, particularly when local networks were connected to one another after 1900, resulting in an interurban transport

Table 5.2 Transit speed and availability of suburban land

Mode	Average speed (mph)	½ hour travel radius (miles)	Land accessible for settlement (square miles)
Walking	2	1	7.07
Omnibus	4	2	12.26
Horsecar	6	3	28.26
Cable car	10	5	78.50
Electric tram	12	6	113.86

Source: McShane (1994: 28).

network (Nye, 1990: 121). Fifth, electric trams stimulated a change in the perception of the function of streets. Before 1890 many streets functioned as a social meeting place, but with the coming of the electric tram, the function was increasingly defined as a transport artery. Another cultural change was the experience of speed. Because of the trolley, city residents became accustomed to higher-speed vehicles in the street. These aspects formed new elements in the transport regime, to which the automobile later linked up.

The bicycle was another transport technology which became very popular in the 1890s. In 1885 the safety bicycle was developed, based on two wheels of the same size and a tubular frame. This new bicycle was more stable, safe and provided easy riding, something which appealed to more user groups. The safety bicycle was technically more complex than the Ordinary, because it used a chain and two sprockets for its driving mechanism. It also included gear wheels, which gave bicycles greater hill-climbing capability. Pneumatic tyres were added in 1888, making cycling more comfortable, and tubular frames were developed to reduce the weight. Chain-drive and gearing mechanisms, tubular frames and pneumatic tyres were new technologies that were later used in gasoline automobiles.

The first safety bicycle was introduced in the United States in 1887, and the Pope Manufacturing Company began production in America in 1889. In the 1890s the bicycle became linked to the new function of 'touring' and to the cultural values of recreation and fun as well as health and 'practising hygiene'. The safety bicycle diffused quickly, leading to the bicycle craze of 1895–96. By that time, the American bicycle industry consisted of more than 300 companies and produced well over a million bicycles per year (Hounshell, 1984: 192). In 1897, the world market for bicycles collapsed, and overproduction brought down the price, bringing it in reach of the working classes. The bicycle became a general work tool, for example, delivery tricycles for freight transport. Another effect of the market crisis was that many bicycle producers diversified their activities to developing automobiles (Mom, 1997: 95). In the 1910s, bicycle use in America declined and bicycles were considered primarily as children's toys.

As the bicycle diffused, it had wider transformational effects. In retrospect, some of the new elements paved the way for the automobile. First, experience with the bicycle gave rise to the articulation of new user preferences and mobility practices. While the horse tram was a public transport mode along fixed routes, the bicycle was a private one, which allowed travel in any direction. Individual and flexible driving were two new functionalities, which were enjoyed and celebrated in the mobility practice of touring. In retrospect, these new functionalities and practices helped to pave the way for the automobile. Looking back from 1937, Hiram Percy Maxim, one of the American automobile pioneers, described the importance of the bicycle in the take-off of the automobile as follows:

> 'We have had the steam engine for over a century. We could have built steam vehicles in 1880, or indeed 1870. But we did not. We waited until 1895. . . . The reason was that the bicycle had not yet come in numbers and had not directed men's minds to the possibilities of independent, long-distance travel over the ordinary highway. . . . We thought the railway was good enough. The bicycle created a new demand which it was beyond the ability of the railroads to supply. Then it came about that the bicycle could not satisfy the demand which it had created. A mechanically-propelled vehicle was wanted instead of a foot-propelled one, and we now know that the automobile was the answer.' (Maxim, quoted in Jamison, 1970: 39–40; Hall, 1998: 398)

Another new element was the creation of bicycle clubs, which developed into lobbying organisations for better roads with smoother surfaces (the Good Roads movement). These clubs set up bicycle journals in which user-related aspects were discussed, for example, the need for signposts and better roads, interesting and scenic touring routes, or a comparison of different kinds of bicycles.

Third, bicycles gave rise to new traffic regulations. The introduction of bicycles made urban street traffic more heterogeneous and dense. Bicycles travelled faster and more quietly than horse-drawn carriages and could easily hit pedestrians (Baldwin, 1999: 216). To facilitate the integration of bicycles, new traffic rules were formulated. In 1897, New York passed a comprehensive traffic code, the first in the nation. A speed limit of 8 miles per hour was established for bikes and other vehicles. Cyclists were required to keep to the right and give hand signals before turning (McShane, 1994: 51).

Fourth, the safety bicycle stimulated the development of new technologies, for example, chain-drive mechanisms (chains, sprockets, gears), tubular frames and pneumatic tyres. These new technologies were later used in early automobiles.

Fifth, the mass production of bicycles stimulated the development of new production technologies. A wide range of special-purpose machine tools were developed and further refined, for example, turret lathes, and milling, screw, grinding, drilling and boring machines (Hounshell, 1984). Through its extensive use of specialised machine tools and metalworking machinery, the bicycle industry pushed further the system of interchangeable parts begun in armouries. But the bicycle industry also introduced new elements in metalworking, for example, sheet metal stamping and electric resistance welding (Flink, 1990: 5). These techniques greatly facilitated mass production and were later used in the production of automobiles. Because of the technical similarities of bicycles and early automobiles, many bicycle producers could diversify into automobile production when the bicycle market collapsed in 1897.

Another important development was the change in road surfaces. Dirt and macadam roads became increasingly problematic, because heavy wagons tore them apart. The 1890s were characterised by a new wave of street improvements, using concrete and asphalt. These improvements took place in the context of the City Beautiful, the Parkway and the Good Roads movements. Concrete had the advantage of low cost and an ability to withstand great weight. But it was also brittle and shattered under repeated blows from iron horseshoes. Hence, asphalt was often added to the concrete, because it provided a hard-wearing surface (McShane, 1994: 61). Concrete and asphalt had smooth surfaces and low rolling resistance, but provided little purchase for horses (Mom et al., 1997: 244). Asphalt was an improvement not so much for horse-driven carriages as for cyclists, street sweepers and city cleaning departments. Street cleaners could operate sweeping machines easily on asphalt roads (McShane, 1994: 79).

The street improvements involved a power struggle between local residents on the one hand and middle-class reformers and public authorities on the other. Did the street belong to those who lived on it or those who used it to get across town? Local residents resisted road improvements, because they saw roads as social meeting places rather than transport arteries. But in the context of wider political and bureaucratic changes, power from local residents was gradually eroded and transferred to public authorities. Middle-class reformers increasingly acquired positions such as municipal engineers, city planners, reform-minded politicians and judges. Thus, they were able to implement legal changes in the administration of streets, which enabled them to push through their ideas about the function of streets. Municipal engineers who took over public works administration, defined streets exclusively as arteries for traffic. Cities also began to allocate a portion of the paving costs from general tax funds, an indication that they planned streets for general traffic rather than for local residents. When local residents took their case to court, they found that the judiciary increasingly favoured commuters over local residents on ordinary streets (ibid.: 117).

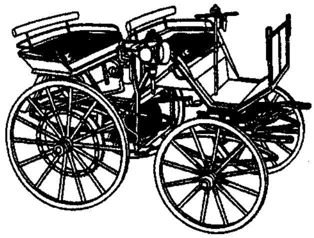
In sum, three changes had occurred. First, by 1900, American cities had built hundreds of miles of parkways and boulevards, paved with macadam, concrete or asphalt. Second, there had been a shift in legal power regarding the administration of streets, which made expansion of road infrastructures in subsequent decades easier. Third, a shift had occurred in the definition of streets: they were increasingly seen as transportation arteries rather than as spaces for social interaction. Thus streets were increasingly the domain of road vehicles, often travelling at higher speeds than was acceptable in previous decades. The automobile would later build upon these changes.

In this fluid context, automobiles emerged as a new technology. Early cars were vulnerable and did not have much horsepower. Since only cities had large stretches of smooth asphalt and brick roads that provided a good context for them, it was initially thought that the automobile would be an urban car. The newly created smooth roads and boulevards thus functioned as proving grounds for early cars. Early developments took place in Europe, but America soon took up the challenge when automobiles were introduced in 1893.

Early automobiles were constructed as an add-on, that is, adding engines and motors to existing coaches or tricycles. This led to several technical trajectories. Electric vehicles had been developed since the 1830s, but batteries and electric motors were heavy and fragile. The lighter batteries by Raimond Louis Gaston Planté (1860) and Camille Faure (1881) stimulated new experiments. Electric vehicles, using lighter batteries, came in three forms: light and heavy coaches, and tricycles. The French coachbuilder Charles Jeantaud added an electromotor to a light carriage in 1881. But heavy batteries remained a problem for these light carriages. Because this problem was less pressing for heavy carriages, they were better suited for early electrification. Small electric motors could be put close to the wheels, and provided with power from a large battery. The existing heavy coach needed little adaptation, giving it a constructional advantage (Mom et al., 1997: 215). Batteries and battery chargers were borrowed from stationary electric applications. Electric motors were borrowed from the electric tram, as was the mechanical controller, a device which allowed easy gearing. There were also proposals for and experiments with electric tricycles, but this trajectory was not viable because there was no clear usage and the heavy batteries were a problem. Electric cars were easy to start and operate at low speeds; they were also easy to accelerate and slow down; and they were clean, quiet, reliable and easy to handle. The electric vehicle was perceived as suited for dignified urban use. Moreover, the electric car had a limited range, which also stimulated urban use.

Light steam automobiles were developed in the 1870s and 1880s in America. To make the handling of steamers easier, petroleum was introduced instead of coal to heat the boilers. Another innovation was Leon Serpollet's flash boiler, developed in 1889, making it possible to start steamers within five minutes. Early steamers used much water, which made water refills necessary every six to eight miles (McShane, 1994: 96). This range was gradually extended to 30 miles in later steamers.

With regard to gasoline cars, two trajectories were followed. Gasoline engines were added either to bicycles and tricycles or to carriages. In



Source: Georgano (1985: 10).

Figure 5.5 Daimler's horseless carriage from 1885

1885, Benz followed the first route, while Daimler chose the second (see Figure 5.5).

The use of the internal combustion engine (ICE) required several technical problems to be solved. Liquid fuels had to be evaporated before entering the combustion chamber. In response to this issue a carburettor was developed. Second, an ICE cannot deliver power from a standstill, because energy is required to start the movement of pistons, connecting rod and crankshaft. Therefore the engine needs to be uncoupled from the wheels when starting the car, which is why a clutch is needed. Third, a gearing mechanism was needed, because the ICE has a small range of optimal number of revolutions. A fourth technical issue was the transmission of power from the engine to the wheels. Several options were tried, for example, cogwheels, belts, and the chain drive borrowed from the bicycle. In sum, gasoline cars differed from electric vehicles in the sense that they needed complex gearing and transmission techniques. Because these techniques required extra space, builders of gasoline cars began to change the chassis of the carriage, creating an extra 'nose' (hood). Over time this 'nose' became longer, while the boxy coachwork was reduced in height. Gasoline cars thus developed a particular form, while that of electric vehicles remained close to existing carriages.

Early gasoline cars had some particular operating characteristics. A common complaint was the difficulty of driving ICE vehicles at low speeds (Mom et al., 1997: 217). The reason was that the ICE works better at a higher number of revolutions. At speeds below 10 km/hour the engine could easily stall. Another characteristic was that the gasoline car was difficult to start: the use of a crank handle required a lot of force. Both characteristics meant that the gasoline car was initially not very suited to use in cities, with slow driving and frequent stops and starts. Gasoline cars were better for driving outside cities and at higher speeds. Since the gasoline car was seen as an 'adventure' machine (see below), it was not a major problem that the internal engine frequently broke down. It required technical skills to repair the engine, something on which the adventurous driver prided himself. But for the same reason, the gasoline car held little appeal to fashionable and distinguished bourgeoisie. Gasoline cars also made a lot of noise, which made driving for pleasure unattractive – but speed devils relished the noise.

In the 1880s much development work was done on different automobiles by automobile pioneers, who mainly aimed to demonstrate technical feasibility. Some of the early cars were bought by wealthy enthusiasts, interested in novelties and mechanical adventure. In the early 1890s the main technical teething problems in electric, steam and gasoline cars were solved. Pioneers now searched for suitable application domains, and four main niches were explored.

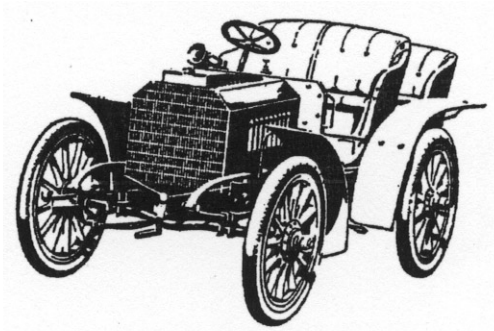
The first was the taxi niche. The gasoline car was not well suited for this, because of the difficulty of driving at low speeds and having frequent stops (ibid.: 217). Steam vehicles were also not suited to taxi use, because of their steam and soot exhaust fumes, as well as their noise. But electric vehicles were easy to start and accelerate. Furthermore, the electric car could link up with a landscape development, the cultural popularity of electricity. With the rapid expansion of the electric tram, it seemed logical that electricity would be the best way to mechanise urban transport. It was thought that the electric car would fill the gap in the electric tram network:

> If, as many believed, the 20th century was destined to be the electrical age, then there was no place in it for the noisy, exhaust-spewing internal combustion engine. . . . The electric car appeared to have all the good points of the horse and buggy with none of its drawbacks. It was noiseless, odourless, and very easy to start and drive. No other motor vehicle could match its comfort and cleanliness or its simplicity of construction and ease of maintenance. (Basalla, 1988: 198)

The limited range of electric vehicles was not perceived as a major problem and could be dealt with through fleets of urban taxis. In the late 1890s many countries experienced a 'lead cab fever', as electric taxis appeared on the streets of major cities: London (1897), Paris (1898) and Berlin (1899). In America, electric cabs were first operated by the Electric Vehicle Company (EVC) in 1899. Between 1899 and 1902, electric vehicles enjoyed great popularity in the taxi niche (Kirsch, 1996: 19). Before racing and touring became popular, it was widely thought that the new automobile would be an urban car. But the enthusiasm for electric vehicles died away, because batteries proved to be vulnerable, suffering from frequent breakdowns. Furthermore, the EVC became embroiled in legal and financial scandals, and went bankrupt in 1903 (Fogelberg, 1997).

The second niche was the luxury one of pleasure riding. Parks were popular places for picnics, tea parties and driving around. Because early gasoline cars were difficult to start and operate at low speeds, electric vehicles seemed an appropriate choice. For this niche, electric 'society cars' were developed, which were exclusive and meant to show off wealth. The limited range of electric cars was not a major problem. In both the taxi and the pleasure riding niches, electric vehicles competed more with horses than with gasoline cars. Manufacturers of electric cars copied fashionable and exclusive carriage forms to tap into an already established market. In these, niches, electric cars bore a close resemblance to horse-drawn vehicles (Flink, 1990: 10).

The third niche was racing. Car races were very popular, attracting many spectators and much press interest. Newspapers printed spectacular photos of car races (Mom, 1997). Car races resonated with wider landscape developments, for instance, cultural values such as excitement and danger, courage, physical prowess, elite status, aggressiveness and striving for records. The first American race was the *Times–Herald* race in 1895. Automobile clubs were set up to organise races in France (1895), Belgium (1896), Italy and England (1897) and the United States (1899). In short-distance races on tracks, electric vehicles were worthy competitors to gasoline cars, taking advantage of their instantaneous energy outburst. In 1899 an electric car set the speed record at 105 km/hour. But in long-distance races electric vehicles encountered problems with the limited energy content of the lead acid battery, and low reliability and sturdiness. Furthermore, batteries took a long time to recharge. A short charge (enough for 20 miles) took 1½ hours, and a full charge could take as long as six hours (Kirsch, 1996: 53). Yet another problem was that electric vehicles, which were the heaviest cars because of their batteries, suffered from kinetic shocks on the rough rural roads. This caused many punctured tyres or damage to the batteries. Gasoline cars fared very well in long-distance races: they had good staying powers, and could easily be repaired when they broke down. The racing niche played a major role in forming the perception of what 'the automobile' should be able to do. Long range and high speed were established as prominent performance characteristics for 'the automobile'. In the context of these criteria the limited range of electric cars was increasingly seen as a problem and an 'inherent' weakness (Abt, 1998). In the racing niche the technical form of gasoline cars changed from a high boxy carriage to a low and long vehicle, more aerodynamic and sportive. When gasoline cars developed a new technical form, traditional electric cars came to be seen as old-fashioned and not in line with modern trends. The new performance characteristics materialised in a first dominant design, the Mercedes of 1901 (Figure 5.6). The Mercedes was low,



Source: Bürgle and Frankenberg (1964: 15).

Figure 5.6 The Mercedes from 1901

long and stable, making it less likely to flip over in races, and it could reach speeds of 53 miles per hour (85 km/h) (Hård and Knie, 1994: 146).

The fourth niche was touring in the countryside. This became very popular in America in the early twentieth century, once cars had become somewhat more reliable. Touring was very popular because it linked up with a range of cultural values at the landscape level. It was associated with curiosity, the exploration of frontiers. It involved aspects of danger and adventure, and the technical skill to deal with them, for example, repairing breakdowns or extricating the car from the mud. Touring was also experienced as 'active travelling', and strengthened two new functionalities that were first introduced with the bicycle: individual and flexible transport. Touring also linked up with the attention being paid to health and nature. Driving cars in the countryside was perceived as 'practising health'. Because touring was a popular niche, different automobiles tried to establish a foothold. There were attempts to use electric cars, but these failed because of their limited range, and the lack of a rural electric infrastructure for recharging. Battery-charging stations were not widespread, except in certain urban areas (for example, around Boston and New York). The electric utilities could have played a stronger role in stimulating an electric-charging infrastructure, but failed to do so. Even if one found a battery-charging station, it could not necessarily be used because battery plugs had not been standardised (Kirsch, 1996: 53), and if it could be used it took some 1½ hours for a short charge. Because of these problems, electric vehicles did not enter the touring niche.

Steam automobiles, however, could build upon an existing infrastructure, watering troughs for horses, to refill their water supply (about every 25–35 miles), and they enjoyed a brief period of popularity and growth (1899–1903). However, in 1914 an epidemic of hoof-and-mouth disease in New England forced the closure of horse watering troughs (ibid.: 50). But the main problem with steam automobiles was that it was unclear which market they targeted (Te Velde, 1997: 10). Steam car manufacturers, who were often master craftsmen wedded to nineteenth century shop practice, did not really champion the use of their cars for touring. The Stanley brothers, for instance, were mainly interested in making beautiful steam cars and cared little for consumers. Market demands were unclear. The Stanley brothers opted to remain a small-scale vehicle manufacturer selling only as many cars as they could comfortably make in their factory (ibid.: 12–13).

Thus touring became mainly the domain of the gasoline car, which was increasingly seen as an 'adventure machine'. The gasoline car could build upon an existing fuel infrastructure. Petrol was widely available at general stores, for example, for lighting purposes and stationary gasoline engines (ibid.: 52). Gasoline cars could also build on an existing repair network consisting of farmers, blacksmiths and mechanics with gas engine competencies. Although touring in the countryside was popular with the upper classes, it met with resistance from the local population. Accidents happened, killing and injuring people or livestock, because drivers travelled at high speeds on bad roads. Speeding cars also caused dust waves (dust problem). In protest, farmers sometimes ploughed roads to make them unusable for cars. The Farmers Institute of Indiana demanded an official ban to keep cars from using rural roads (Rao and Singh, 2001: 253). The opposition from farming communities and anti-speed organisations jeopardised the standing of the automobile. In reaction, automobile clubs lobbied for favourable legislation and defended drivers against criminal suits. Public authorities began a regulatory process to embed the car institutionally, thus defusing opposition (ibid.: 263). Car registration and car tags were introduced to facilitate identification of cars and drivers in the event of accidents. An institutional response to reckless driving was the licensing of drivers. Speed limits were introduced, but these were rather favourable. Local speed limits were raised from 6 to 12 miles per hour within city limits in 1901 and to 20 miles per hour outside the city limits in 1905 (Baldwin, 1999: 216).

In sum, by 1903 a stable distribution of functions had emerged. Electric vehicles were defined as urban cars, used in the taxi and the pleasure riding niches. Gasoline cars were defined as 'adventure machines', used in the racing and touring niches. Because these different niches were relatively separate, electric and gasoline cars did not compete directly. An equal number of steam automobiles and electric and gasoline cars were sold in the early years, but market articulation remained unclear, hampering the diffusion of the 'steamers'.

5.4 EXPANSION OF ELECTRIC TRAMS AND STABILISATION OF AUTOMOBILES (1903–1914)

Landscape Developments

An ongoing process at the landscape level was immigration. Between 1892 and 1920, 12 million migrants passed through New York (O'Rourke and Williamson, 1999: 119). Overcrowding, health problems and social reform continued to be important issues. The Progressive Reform movement continued until the First World War. Progressive reformers saw suburbanisation as a way to deal with urban social problems. Because automobiles could be instrumental in facilitating suburban living, reformers were supportive of mass motoring (Ling, 1990: 6–7). The automobile thus functioned as a political tool to facilitate wider social change.

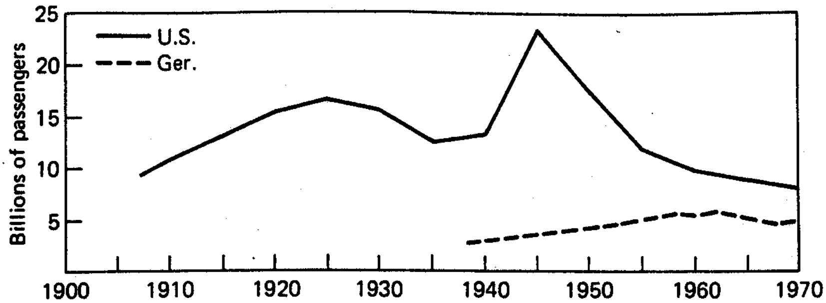
Progressive reformers were concerned not just about the city but also about the state of rural America. Isolation of farm people and the weakening of rural institutions (church, school, shops) were often discussed. Small schools, churches and shops were closed and consolidated into larger ones, because these could be run more efficiently. The automobile came to be seen as a way to achieve consolidation and efficiency (ibid.: 22).

The embrace of the automobile as a means for social change fitted well with another landscape development, the movement of technological utopianism between 1880 and 1930 (Segal, 1986). Technological utopianism derived from the belief in technology as the means of achieving a perfect society in the near future. The dirt, noise and chaos that characterised American industrial cities were to give way in the future to perfect cleanliness, efficiency, quiet and harmony.

Technology and engineers could acquire much esteem in a context of concerns about corruption of public authorities (the 'good governance movement'). It was thought that professionals, experts and engineers would do a better job than politicians of finding rational solutions that could satisfy everyone. In this context municipal engineers, city planners and highway engineers could acquire powerful positions in administrative agencies and public bodies.

Expansion and Problems in the Electric Tram Regime

The electric tram regime enjoyed a great expansion in the first decades of the twentieth century. By 1900 there were about 850 electric tram systems in American cities operating over 10,000 miles of track. The electric tram became the dominant urban transport mode. Figure 5.7 shows the rise in yearly passengers.



Source: Yago (1984: 9); courtesy of Cambridge University Press.

Figure 5.7 Passengers in the electric tram in the USA and Germany (per annum, bn)

The electric tram was widely used by the middle classes for commuting between work in the city and suburban houses. Transit was essential for the white middle classes in America to enjoy a decent life in the suburbs. The strongest growth occurred in the interurban tram systems, operating between cities. The electric interurban mileage increased from 2107 miles in 1900 to a peak of 15,580 miles in 1916 (Flink, 1990: 3). This system opened up the city for rural residents for amusement, schools, buying and selling produce (Nye, 1990: 119). It also opened up the countryside to urban residents, for instance in the form of tourism, day trips and picnics, and it opened up new land for residential areas. Trolley companies, real-estate developers and land speculators could make huge profits from real-estate speculation and increasing land values (Hall, 1998: 809).

Despite its expansion, there were several problems with the electric tram regime. First, the trolley was strongly regulated with regard to fares and transfers. City councils and state commissions refused to allow tram companies to increase the price, which for a long time remained at five cents (Nye, 1990: 134). They also instated universal transfer, paying per ride rather than according to distance travelled. These regulations made it difficult for tram companies to increase their income (Flink, 1990: 363). A second problem was the declining standard of comfort, in particular overcrowding at rush hours, which resulted in delayed services, breakdowns and dirty carriages. Another reason for complaint was the inflexibility of routing. As factories were set up on the outskirts of cities, workers often first had to travel to the city centre, and then take another tram to the factory. Users thus became dissatisfied with the service offered by electric trams. A third problem was that tram companies acquired a negative public image. In the 1890s and the early twentieth century they had enjoyed large profits, but the public felt that these had not been invested to improve the service. Trolley companies seemed to be more interested in money than in 'proper' transportation. Furthermore, some trolley companies were involved in real-estate speculation, and publication of speculative excesses gave them the public image of greedy and rich monopolists. Hence the public had little sympathy with the increasing financial problems in the late 1910s and 1920s, believing that trolley companies had hidden reserves (Ling, 1990: 73). During the First World War, high costs for steel rails, coal and equipment worsened the trolley's financial situation. The traditional five-cent fare was not enough to support the transit industry, let alone fund expansion and improvement. But public authorities were disinclined to help, because of the negative public image of tram companies.

Stabilisation of Gasoline Cars (1903–1914)

In 1900, automobiles were still toys for the upper classes, playing a minor role in terms of overall mobility. Gasoline cars were used by rich sportsmen and adventurous tour drivers, while electric vehicles served as luxury cars in cities (for example, promenading, tea parties). From 1903 to 1907 this relative separation persisted: steamers and electric vehicles maintained their share in their market niches (see Table 5.3), but the number of gasoline automobiles raced ahead, because the application domains of car racing and touring enjoyed great popularity. By 1905, the market was tipping decidedly towards internal combustion. The cultural popularity of the racing and touring niches led to closure regarding the characteristics of 'the car' (Kirsch, 1996; Fogelberg, 1997: 37; Mom, 1997: 124): it should have great power, high speed and long range.

The increasing sales of gasoline cars provided resources and incentives for engineers to further refine the components of the car. Many technical

Table 5.3 Annual car sales in the United States, 1900–1905

	1900	1905
Electric cars	1,575	1,425
Steamers	1,681	1,568
Gasoline cars	936	18,699
Total	4,192	21,692

Source: Kirsch (1996: 60–61); Mom (1997: 223).

components had to be improved to make it stronger and more reliable. To improve the power of the gasoline engine, the size and configuration of cylinders, and the placement of valves and camshafts, were examined. The flywheel magneto (1908) provided an electrical power source for firing the spark plugs that were built into the engine. The mounting of the engine in the car and the architecture of components were also topics that received much attention: 'From 1903 to 1907, the spark advance, the longitudinal mounting of engines, torque tubes for drive shafts and bevel gear systems were introduced in successive Ford models' (Abernathy and Clark, 1985: 8–9). The gasoline car increasingly deviated from traditional carriages, as the community of car engineers developed new design rules. An important innovation was the vanadium steel alloy (1908), stronger and lighter than traditional materials. This alloy could be used in engines and chassis components, and was important in the development of lightweight vehicles, which were strong and reliable. The accumulation of component innovations improved the functioning of the gasoline car.

After 1905, new groups began using cars for business purposes, for example, doctors, travelling salesmen and insurance agents. Wealthy farmers began to use cars as a practical transport vehicle to circumvent the train's monopoly (Hård and Knie, 1994: 147). The rural opposition to cars weakened as wealthy farmers switched sides. After 1908, rural newspapers stopped calling cars 'devil wagons' (McShane, 1994: 178). Another new commercial niche was that of the taxi. Gasoline taxis appeared in New York in 1907, replacing horse-drawn hackneys. The taxi niche was the trailblazer of a more utilitarian use of gasoline cars. Gasoline cars thus began to encroach upon market niches that were previously the domain of electric vehicles. The use of cars for touring remained important, but gradually changed its character: after 1910 it developed into autocamping, a vacation alternative for the middle classes (Belasco, 1997: 7) who carried their own equipment for repairs, setting up camp, making fire and cooking meals.

In the early twentieth century, a new design trajectory of cheap, strong and sturdy cars emerged, pioneered by people such as Ransom Olds and Henry Ford. At the time, automobiles were still expensive products, large and luxurious. The innovation of Olds and Ford was that they had a vision in which the car was not a luxury for the few, but a basic convenience for many. After years of experimenting, Ford found the right balance between cheap price and sufficient robustness in his Model T, priced at $850 in 1908. Its basic merit lay in the combination of lightness, sturdiness, simplicity and power. The Model T was a success, and sales rose rapidly. Around the same time the luxury market became saturated, indicated by a sales drop in luxury cars in 1907. The sales crisis was a turning point with regard to the strategies of car manufacturers (Mom, 1997). To increase their sales, car manufacturers had to find new markets and thus they embraced the idea of going downmarket. The importance of the Model T was that it created a dominant design for the car industry. Although the Model T originally existed in a variety of body styles (for example, touring car, open runabout, town car, torpedo runabout, roadster) Ford decided in 1909 to build only one model, with the same chassis and in one colour. It was rugged and reliable, and had a maximum speed of 45 miles per hour. Ford deliberately gave his car high axles and three-and-a-half-inch-wide tyres, so that it could better traverse roads deeply rutted by farm wagons, and consequently rural areas formed a major market niche for the diffusion of the Model T. In the 1910s, farmers bought cars, because it provided a partial answer to rural problems (isolation, declining rural institutions).

With the Model T as the dominant design, further innovations could proceed 'down the design hierarchy' (Clark, 1985), focusing on further improvements in subsystems and components. With regard to engines, the number and configuration of cylinders and the starting/firing of the engine received much attention. An important innovation was the electric starter, developed in 1911 by Charles Kettering. Crank starting had been a major problem, because it required great force and dexterity. By using high-voltage ignition on the basis of batteries, the electric starter solved the problem of starting gasoline cars. The gasoline car was thus improved by borrowing technical elements from the electric vehicle, increasingly its competitor.

As gasoline cars became easier to start and operate, by 1910 they increasingly invaded urban market niches of electric vehicles. Electric vehicles had maintained some luxury market niches in urban transport, for example, driving for pleasure and technically unskilled drivers (elderly people, women). Electric vehicles had maintained these niches, because of their convenient operating characteristics (easy starting, clean, reliable). But as gasoline cars improved, electric vehicles lost their comparative advantage. The former increasingly invaded the urban luxury niches of electric vehicles.

The number of cars rose rapidly in this period, from over 21,000 in 1905 to 45,000 in 1907 and 485,000 in 1913 (see Table 5.1). But the use of the electric tram, still the dominant urban transport system, also increased in this period (Figure 5.7). Although the car was less important in overall mobility than the tram, its rapid diffusion was accompanied by the articulation of a range of other elements. One was the improvement of streets and infrastructure. Not all streets were improved with similar speed. On radial avenues and main thoroughfares, asphalt paving was implemented in the 1910s. On side streets, however, water-bound macadam, gravel and dirt surfaces remained common (Baldwin, 1999: 206). Important streets with much traffic were widened, usually by cutting back on sidewalks or cutting down trees. Local residents objected, but the demands of suburban commuters and shoppers took priority.

Another new element was the institutional embedding of the car and the formulation of a new set of formal traffic rules. Traffic regulations in the nineteenth century had given priority to the safety of street users rather than to rapid traffic flow. Pedestrians, bicycles, trolleys, carriages and wagons had mingled on urban streets. Safety was enforced with low speed limits in some places, around 6–8 miles per hour (ibid.: 216). But with rising numbers of electric trams, cars, bicycles and freight wagons problems increased, for example, congestion, accidents and chaos. New traffic rules were devised to create more order on the street, but in the early twentieth century these traffic rules were organised around a new guiding principle: efficient flow and high vehicular speed became more important than pedestrian safety. Automobile owners, state legislators, city officials, police and educators worked together in the articulation of new rules. They cleared the way for high-speed travel by relaxing the speed limits, adopting regulations to ease traffic flow, and undertaking educational efforts to control pedestrian behaviour (ibid.: 214). Local speed limits were raised to 12 miles per hour within city limits in 1901 and to 20 miles per hour outside the city limits in 1905. In 1903 the New York police commissioner issued traffic rules to separate different traffic modes to facilitate the flow (McShane, 1994: 185). Slow vehicles (carts, bicycles) should stay on the near side of the street, fast vehicles in the middle (automobiles).

Yet another aspect of an emerging socio-technical regime was the articulation of new user routines and mobility practices. This was related to the increasing number of accidents with automobiles. The automobile was perceived as a dangerous transport mode, because of its high speed and impact. In New York, the number of fatal accidents caused by cars grew rapidly after 1910 (ibid.: 175). About 75 per cent of the fatalities were pedestrians, often children playing in the street. In response to these accidents, several actions were taken. One was to build more playgrounds to give children alternative spaces for playing, and 'isles of safety' were built to protect adult pedestrians (Baldwin, 1999: 220). Another was to teach adults and children new routines. In the 1910s, educational campaigns at public schools taught that pedestrians should remain alert and follow special procedures when crossing the street. Public opinion and newspapers played a role in disciplining pedestrians. Before 1905, New York newspapers mostly criticised car drivers for speeding, dangerous driving and causing accidents. By 1907 this had changed, and the *New York Times* began urging mothers to keep their children off the streets (McShane, 1994). Pedestrians were blamed for their own carelessness if they were hit by a car. Many drivers came to expect pedestrians to get out of the way and developed the habit of honking instead of slowing down (Baldwin, 1999: 223). Another response to accidents was the expansion of driving schools, to teach drivers new routines. The task was not so much to teach general driving skills, but to discipline the drivers not to speed, not to get agitated and to respect general traffic rules. Before the First World War a 'morality of good traffic manners' was articulated (Mom et al., 1997: 174).

While cities experienced the early emergence of a new socio-technical regime around cars (road improvements, traffic rules, pedestrian routines, driving practices), things moved somewhat more slowly in rural areas. The main reason was that American rural roads were still in a bad state, consisting mainly of dirt roads. While $300 million was spent on the improvement of urban roads in 1906, only $75 million was spent on rural roads (McShane, 1994: 218). In the 1910s, several social groups began to lobby for improvements of the rural infrastructure, including good-road campaigners and farmers, hoping to break the railroad's transport monopoly. Road builders and cement manufacturers also lobbied for rural highway improvement, as did motorists, organised by the American Automobile Association. In reaction to this pressure new public departments were created, staffed with highway engineers. Because corruption was endemic in public works projects, highway engineers earned great respect and power because they were seen as honest and rational. Highway engineers seized the opportunity to establish themselves as a professional body, and via the American Association for Highway Improvement (AAHI), founded in 1911, they claimed to speak with scientific authority on matters of highway policy (Ling, 1990: 53). Although rural road expansion was still limited in the 1910s, an institutional and organisational framework was created, invested with trust and prestige with regard to road building. This organisational framework laid the foundations for the road-building boom of the 1920s and 1930s.

Niches: Electric Vehicles and Motorbuses

Electric and steam cars remained viable in some niches for several years after 1910, but formed no threat to the dominance of the gasoline car.

By 1905 a second generation of electric cars, characterised by more reliable and stronger batteries, was exhibited at car shows. (Mom, 1997: 224). These new electric vehicles were more reliable and could attain higher speeds. However, the electric vehicle did not stabalise in a dominant design, thus preventing mass production, scale economies and lower cost. Between 1907 and 1910 there was a brief resurgence of interest, the 'golden age' of electric vehicles (Wakefield, 1994). But the electric car was increasingly confined to small market niches, becoming a 'society vehicle' for tea parties, promenading, or going to the theatre or tennis club. In the domain of freight transport, however, electric trucks found a market niche, where they could hold on for several decades (Mom and Kirsch, 2001). In 1913, for instance, electric cars accounted for about 22 per cent (almost 3900 cars) of business and commercial cars (Mom et al., 1997: 179). Nevertheless, despite the revival of the electric car it was clear that the electric vehicle had lost the battle with the gasoline car. After 1911, gasoline cars with electric starters increasingly invaded the urban niche of electric vehicles.

A new technology was the motorbus. In 1907, New York was the first American city to purchase a bus with an internal combustion engine. Mechanical teething problems were solved in the mid-1910s, and by 1920 there were 500 buses in New York. However, no American city other than New York had more than 30 buses in operation by 1920. Buses initially diffused slowly, because bus firms sought an elite ridership, charging ten cents a ride, twice the trolley fare (McShane, 1994: 192).

5.5 TOWARDS A CAR-BASED PERSONAL TRANSPORTATION REGIME AND ITS IMPACT (1914–1930S)

The De-alignment of the Electric Tram Regime

The regime of the electric tram gradually disintegrated in the 1920s and 1930s, as a result of de-alignment on multiple dimensions, for example, regulation, finance, public perception, management practices, public policy and user preferences. The electric tram was strictly regulated. During the 1910s, city authorities used municipal franchises to force tram companies to stick to fixed fares of a nickel, even though costs increased for tram companies (wages, material costs). Unprofitable operations could not be terminated and fares could not be raised to levels that ensured a reasonable return on invested capital (Flink, 1990: 362). Only in the late 1920s were fares allowed to go up from a nickel to seven cents. As a result, financial problems increased for tram companies. However, the public had little sympathy with their financial problems, because tram companies had made windfall profits in the early years, and had acquired a public image of being greedy and rich monopolists (Nye, 1990: 135). This also had a negative influence on their relationship with local authorities. Tram companies had to pay heavy taxes to city authorities. While the trolley was taxed, the private passenger car and motorbus were massively subsidised by publicly funded street improvements to accommodate automobile traffic (Flink, 1990: 363). Local public policy makers embraced the car and the motorbus, while they fought with trolley companies. In cities throughout the United

States, street railways faced antagonistic urban officials. Politicians saw transit companies as convenient villains. Politicians often opted for the motorbus as part of a political power struggle (Schrag, 2000):

> Their [that is, transit executives, politicians] decisions to abandon streetcars, in New York and in other cities throughout America, cannot be explained either by inherent technical advantages of buses or the conspiracies of bus manufacturers. Rather, the bitter antagonism between tram transit companies and local politicians moved both the companies and the politicians toward support of the bus as a means to rewrite old rules. (p. 52)

During the 1920s, the undercapitalised trolley companies made few investments. Tram companies were burdened with overextended lines, declining investment, rigid fare structures and rising costs for replacing equipment (Nye, 1990: 135). The financial problems were made worse by poor strategic management. The trolley was ill-prepared for the increasing competition from buses, trucks and automobiles which offered more flexible routes, faster service and greater independence of movement (ibid.: 134).

User satisfaction with the trolley also declined after the mid-1910s. Travellers wanted a safe, smooth and temperature-controlled ride, frequent and punctual service, comfortable and assured seating in a clean vehicle, secure shelter while waiting, adequate night 'owl' and weekend service, and routes that required a minimum of walking and transferring (Davis, 1995: 311). On many of these criteria the trolley services were failing. Because of the lack of investment, trams became more crowded, especially during the rush hour. Another effect of low investment was a rise in the frequency of breakdowns and a decrease in punctuality. There were also more qualitative and cultural issues in the choice between bus and tram. One issue was America's desire for novelty. Because trams had high capital costs, they depreciated slowly. Trolley coaches could have a life expectancy of 12–15 years. By contrast, early gasoline buses generally lasted only three years in urban service. Thus, while tram companies stuck to old trolleys, bus companies could improve continually and use the latest bus models. Buses came to be seen as young, modern and vital, while trolleys were seen as old technologies of the past. Another issue was that the *experience* of speed changed. Automobiles revolutionised ideas of speed. Motorcars moved through traffic some 40–45 per cent faster than the average tram (ibid.: 317). Another problem was safety. Sexual harassment was a continual complaint from women. The overcrowding in trams created ample opportunities for 'accidental' contact and blatant fondling. Safety problems were mixed with those of racism. Transit officials received complaints about unclean clothing of Negro workers on the car lines (Ling, 1990: 86).

In sum, the de-alignment of the electric tram regime was the result of a range of political, social, economic and technological causes. The number of passengers began to decline in 1924 (Hilton, 1969: 128–9). Rapid decline of electric mass transit took place in the 1930s. In the mid-1930s the tram industry made an effort to modify its technology to meet competitive pressures. In 1935 the PCC car was introduced, a high-speed tram car of high capacity for the main routes, which were still considered viable (ibid.: 129). The effort, however, was too late. During the 1930s and 1940s trolleys gradually disappeared from American cities (Nye, 1990: 136). The decline in trolley track mileage levelled off temporarily during the Second World War. But after the war, tram ridership fell sharply and consistently. In 1931 about 67 per cent (9.2 billion passengers) rode by streetcar, compared with 16 per cent (2.3 billion passengers) by motorbus. In 1953, 14 per cent (or 2 billion passengers) rode by streetcar; 60 per cent (8 billion passengers) by motorbus, plus another 1.4 billion by trolley bus (Rae, 1968b: 305–6). By 1955 all modes of public transit were in decline. The culprit was the private passenger car, which in the 1950s began making dramatic inroads into ridership on all modes of public transit (Flink, 1990: 367).

The Rise of the Automobile Regime

Elements of a socio-technical regime based around the automobile were already being put in place before the First World War, for example, institutional embedding of the automobile, expansion of urban road infrastructures, formulation of traffic rules and mobility routines. In the 1910s and 1920s more elements were articulated and linked together in an increasingly robust socio-technical regime. Major investments were made in production technologies (Ford's assembly line) and road infrastructures. Public authorities chose to encourage the automobile and neglect the electric tram. Between 1914 and 1930 the automobile was sold in ever-larger numbers. Its ever-wider diffusion had two complementary drivers: first, a major cost reduction, enabled by mass production; and second, the creation of a new socio-technical regime, consisting of new infrastructure, traffic rules, behavioural routines and mobility practices. Both of these drivers are analysed in this section.

The first driver was the articulation of a new production system, mass production and the resulting cost decrease. This new production system was pioneered at the Ford Motor Company. Even before 1910, Ford realised that mass production was the way to produce cheaper cars. To raise productivity, Ford engineers turned their attention to production flow, focusing particularly on materials handling (Biggs, 1996: 108). In 1910, Ford built a factory at Highland Park, which provided the locus for many experiments with improved production flow. Ford allowed his engineers to carry out extensive experimentation in the factory, building upon prior developments such as: production of interchangeable parts, the idea of a continuous moving belt or conveyor line, strong subdivision of labour, the use of single-function machine tools and sequential ordering of machines depending on the sequence of work (Nye, 1990: 223–5). The creative element introduced by Ford was that these work practices were combined into a functioning whole. On 1 April 1913, real-life experiments began with the first moving assembly line for manufacturing a part of the car, the flywheel magneto. No longer did the men stand at individual workbenches, each putting together an entire flywheel magneto assembly from the many parts. The workers now had to place one particular part in the assembly or start a few nuts and then push the flywheel down the row to the next worker. The workers repeated the same process over and over again. The experiment was a great success in terms of productivity. While 29 workers had assembled 35 or 40 magnetos per day at the workstations (roughly one every 20 minutes per person), 1188 were produced on the line in the first day (roughly one every 13 minutes per person) (Hounshell, 1984: 248). Within the next year, performance was increased to an output of 1335 flywheel magnetos in an eight-hour workday (five man-minutes compared to the original 20). In sum, productivity gains were enormous. Hence, the principle was applied to other assembly operations, engines, chassis, transmission, magneto coils and so on. The shift to the assembly line at Ford was swift. Within a year, virtually every assembly operation had been put on a moving line basis (ibid.: 247). While Highland Park was the first 'niche' for the assembly line, it reached its full potential at the River Rouge plant, opened in 1920. This massive plant introduced the modern rational factory to the world, combining rational factory planning with modern production and construction technology (Biggs, 1996). Mechanised systems of materials handling, automated machinery and the elimination of belting for transmission of power had freed engineers to lay out the shop floor in the most efficient manner. The Rouge plant represented the direction in which modern industry would move, in both building style and production methods.

Mass production allowed Ford to offer the Model T at increasingly cheaper prices. In 1909–10, he raised the price of the Model T from $850 to $950, including extras, to pay for the Highland Park plant. Then the price dropped to $780 in 1910, $690 in 1911, $600 in 1912, $550 in 1913, $490 in 1914, $440 in 1915 and $360 in 1916. America's entry into the First World War increased the price to $450 in 1918, after which it again decreased (Hall, 1998). Ford continued to make improvements in the assembly line in the 1920s. Reliable electric welding of metal parts in high-volume applications became possible when Ford developed a mechanical seam welder

(1925). Around 1924 new fast-drying paints were developed, which made the mass production of coloured vehicles possible. Other innovations included the development of special-purpose machine tools, go–no-go gauges, improved glassmaking procedures, pneumatic tools, and the development of metal stamping machines for large panels (Volti, 1996: 670). All these improvements helped to reduce the cost.

The strong price decrease of the Model T made the car available to more consumer groups. Between 1913 and 1924 the number of cars grew from 485,000 to over 3 million (see Table 5.1). The demand by farmers created a major market niche in the 1910s. Cars provided an answer to some ongoing rural problems such as isolation, and the decline in rural institutions. The car was popular with farmers because it could bring them to church, to larger shops, and transport their children to school: 'The automobile fulfilled a demand for a social alternative to the dysfunctional communal life of the open countryside by 1910' (Ling, 1990: 20). The car was also popular because farmers could use it to undermine the monopoly of train companies, against whom they held a long-standing grudge. The farmers were innovative users, who saw the Model T as a general power source. By jacking up the rear axle and running a belt over one of the wheels, the car was turned into a general power source which was used for driving mechanical tools (for example, shellers, grinders, saws, water pumps), agricultural tools (hay balers, fodder and ensilage cutters, cider presses), or powering washing machines (Kline and Pinch, 1996: 775). Middle-class suburbanites formed another major market niche in the 1910s. They used their cars for commuting longer distances to their suburban homes. The car greatly stimulated suburban living, and residents also used the car for touring in the countryside, which became a very popular recreational form. In the early 1910s, touring was still free and independent, indicated by the term 'gypsying'. But as the number of autocampers grew into the millions, new institutions emerged. Around 1920 free municipal campgrounds were set up, centrally located in towns along major touring routes. Campsites offered more convenience (for example, running water, shops) and additional facilities such as electric lights, hot showers and central kitchens (Belasco, 1997: 4). In 1925 some private campsites began offering cabins for tourists desiring more comfort and privacy. Thus the motel industry was born. The gypsy gave way to the consumer who valued comfort, service and security.

The wide diffusion of cars was accompanied and made possible by wider changes in the socio-technical regime. The articulation and linking together of new elements was the second driver of wide car use. One element was institutional embedding, something which had begun in the early twentieth century. Speed limits were introduced, but these were not very restrictive and were gradually raised over time. Traffic rules forced slow traffic to the side of the road and created more space for cars. Educational campaigns taught children and pedestrians new routines for crossing the street and playgrounds were created to keep children off the street. As a result of these new regulations, by 1920 the street was the exclusive property of the automobile. On the one hand, this amounted to a monopolisation of public space by the car, but on the other it resulted in a decrease in the number of traffic casualties. Furthermore, the flow of traffic was improved. But these gains were offset by the growing number of cars, leading to congestion and parking problems. Permanent twice-daily traffic jams spread to every large American city in the summers of 1914 and 1915 (McShane, 1994: 194). The first solution to congestion was new regulation to increase traffic flow, for example, banning left-hand turns on major avenues, and 60-minute parking limits. The best regulatory results, however, came from municipal engineers who worked out a system of traffic signals after 1914 (ibid.: 200). These lights on poles were more visible than police officers and could be synchronised to speed up the flow. The second solution to congestion was to expand road infrastructures, something advocated by urban planners, and city engineers. These two groups created a new suburban territory around cars, and reshaped downtown areas by resurfacing, street widening, and bridge construction (Lewis, 1999: 44). The reform efforts also resulted in a new kind of road: the limited access highway to be used solely by motor vehicles. Plans for building highways became more widespread during the 1920s. Urban policy makers chose to invest massively in the improvement and construction of roads, but only modestly in public mass transit (Flink, 1990: 369–70). This policy choice favoured middle-class car owners over working-class streetcar riders.

Public authorities were strongly involved in road building. In 1916, President Woodrow Wilson signed the first Federal-Aid Road Act into law, creating the federal Bureau of Public Roads (BPR) to coordinate technical matters. Highway engineers were given considerable authority to develop technical rules and working practices for highway construction, which became a vital ingredient in the booming economy of the 1920s (Ling, 1990: 59). The BPR developed into a well-functioning 'highway machine', involving a large network of interrelated groups with interests in road building, for example, the Portland Cement Association, the American Automobile Association, the American Road Builders Association, the Association of Highway Officials, the Rubber Association of America, the Mississippi Valley Association of State Highway Officials, the National Paving Brick Manufacturers' Association and the National Automobile Chamber of Commerce (Lewis, 1999: 13). The BPR used booklets, films and visiting speakers in campaigns to convince people about the importance of good roads. The Depression increased federal highway funding as an antidote to unemployment. In 1930, Herbert Hoover raised federal aid from $75 million to $125 million. Between 1933 and 1940, the New Deal resulted in more than $1.8 billion being spent on new roads (ibid.: 23). Under pressure from the highway lobby, the 1944 Federal Aid Highway Act increased the federal contribution to state highway building to 60 per cent. Another important step was taken with the 1956 Interstate Highway Act, whereby the ratio of federal funding was increased to 90:10 (Flink, 1990: 371). An argument for the interstate system was its strategic need for national defence (enabling trucks to transport supplies and materiel). Passage of the 1956 Interstate Highway Act ensured the complete triumph of the automobile over mass transit alternatives and killed off (except in a few large cities) the vestiges of balanced public transportation systems that remained in the 1950s.

With the creation of new infrastructures, new uses for cars opened up in the 1920s, for example, suburban and intercity travel. These new uses were associated with new customer groups, which were less rural, less mechanically sophisticated, less tolerant of inconvenient operation and less wealthy. New designs in suspensions, bodies and transmissions redefined the nature of the automobile (Abernathy and Clark, 1985: 12). Ease of operation, smoothness of ride, comfort, convenience and power became important selection criteria. The concept of the 'rolling living room' became popular. While Ford's innovation strategy focused on lower costs, standardisation and improved assembly-line production, General Motors and Chrysler explored new technical trajectories, developing new designs in suspensions, bodies and transmissions. The innovation that contributed more than any other to this change was the closed steel body (ibid.: 12). The innovation introduced new criteria for automotive design: passenger comfort, room, heating and ventilation. General Motors and Chrysler introduced new concepts in design and marketing. One of the new marketing practices was the enshrinement of styling as the prime means of attracting buyers. Chrysler's strategy of new design concepts was highly successful in the 1930s. By the end of that decade Chrysler had surpassed Ford in market shares (ibid.: 20).

By the 1930s the automobile industry was a huge consumer of semi-finished and finished intermediate products (sheet steel, glass and paint) and components (tyres, lamps, generators and so on). It was the largest consumer of machine tools, offering a rapidly growing market to heavy industry (Landes, 1995: 443). Table 5.4 shows the supply network that fed into the car industry. The industry also posed technical questions to basic disciplines such as metallurgy, organic chemicals and electrical engineering. And the use of the car boosted the petroleum industry as well as construction and public works (roads, bridges, tunnels). In sum, the car had many forward and backward linkages in the American economy of the 1930s.

Table 5.4 Share of automotive consumption in total consumption of selected products in America (%)

Product	1929	1938
Steel (all forms)	18.0	17.0
Strip steel	60.4	51.0
Bar steel	28.7	34.0
Sheet steel	29.2	41.0
Alloy steel	–	54.0
Malleable iron	52.0	53.0
Plate glass	73.0	69.0
Rubber	84.2	80.0
Aluminium	37.4	10.6
Copper	15.7	12.1
Tin	23.6	9.2
Lead	31.2	35.1
Nickel	26.0	29.0

Source: Landes (1995: 443).

Because many elements were linked together into a socio-technical regime, the automobile acquired societal momentum and 'dynamic rigidity'. Staudenmaier (1989) describes this reversal as follows:

> Between 1915 and 1970 America responded to the automobile with a host of institutional creations from motels and suburban architecture to legislation governing highway funding, insurance and licensing. In the process, millions of individuals and hundreds of institutions coalesced in a little-noticed social grouping that can be called the automobile maintenance constituency. What gave them social coherence (despite their wide economic, political and cultural diversity) was the fact that they all had come to benefit from and depend on the increasingly popular automobile system. . . . Ordinary citizens also adopted the automobile as an essential means of transportation. (pp: 154–5)

The very size and diversity of the network of social groups indicates the societal momentum achieved by the automobile. Although the automobile was not yet dominant in terms of market shares in the 1930s, it was widely thought that it represented the way to the future.

The embedding of the automobile in American society occurred in tandem with wider impacts on society. Many of these social and cultural changes began before the arrival of the car, but were greatly accelerated as the automobile diffused in society. In that sense the changes were not simply 'caused' by the car. One impact was the transformation of urban life. In the

1910s and 1920s an urban traffic system was formed, dedicated to an easy flow of automobiles, at the expense of residential use of streets and pedestrian safety. The urban environment was adjusted to give cars dominance on the streets. Public space was segregated, producing separate areas for particular functions, for example, children's playgrounds, a central business district and suburban housing:

> Americans reshaped the environment through which the automobile moved to suit the automobile's requirements. As much as half of a city's land area was dedicated to roads, driveways, parking lots, service stations, and so on. Features that were not friendly to the automobile, such as narrow streets and low clearances, were eliminated, along with most alternative forms of transportation. . . . This reshaping of the environment was not caused by the automobile itself. Americans were extremely active in defining their landscapes by means of zoning boards, park commissions, and city councils. (Nye, 1998: 180)

Although suburbanisation was accelerated by automobiles, it built upon a cultural preference for suburbs, which were envisaged as clean, safe, relaxing and green antidotes to the city. It was also stimulated by public policy. Government programmes subsidised suburbanisation through home mortgage tax deductions, and mortgages from the Variable Housing Allowance program (VHA) and the Federal Housing Administration (FHA) (Volti, 1996: 683–4). A second societal impact was the transformation of rural life. Schools, churches, shops and medical facilities were already weakened in rural areas. The car was promoted as a solution to these changes, but it also further stimulated them. As car ownership spread, it gave politicians further reasons to close small schools, shops and churches (Ling, 1990: 23). Small urban shops closed down, as people with cars went shopping in larger centralised retail outlets. Third, the wide diffusion of the car facilitated the emergence of a 'car culture', supported by institutions such as fast food restaurants on highways, shopping malls on the edge of cities and drive-in movies. Touring with automobiles developed into a popular recreation activity.

The Continued Existence of Horses in the Niche of Freight Transport

Freight transportation formed a market niche, where horse transport could retain a foothold, while it was driven out of other urban applications. The advantage of the petrol engine over the horse was not so obvious for freight as for passengers, especially because the larger fixed capital cost was often offset by much time lost in waiting for, and despatching, loads (Barker, 1993: 160). Freight transportation thus provided a niche where horses and related social groups (for example smiths, wheelwrights, cart, van and wagon makers, fitters, painters, coachmen, carriers, horse-keepers, stable-keepers) could persist. Because horse-drawn transport could continue in this way, the related social groups were not immediately threatened with unemployment, which might cause social unrest:

> The continued use of horses for freight haulage long after they ceased to be used for passengers explains why we hear so little about unemployment resulting from this important transition. Some horse men certainly switched to motors, but the pace of the transition allowed those no longer needed to service or drive horse-drawn passenger vehicles to switch to horse-drawn carts, vans or wagons and gradually retire from the transport business rather than to be thrown suddenly out of work. The number of horse-drawn vehicles in Britain, still over 200,000 in 1923, took the rest of the decade to fall to 50,000. By 1937 there were still nearly 12,000 of them, many owned by railway companies. (Barker, 1993: 160)

5.6 CONCLUSIONS

In this concluding section I shall follow the case-study protocol, described in Chapter 3 (Section 3.5). I shall test first the multi-level perspective and then the different patterns.

Testing the Multi-level Perspective

In Chapter 3 (Section 3.5), two questions were formulated as proxies to test the usefulness of the multi-level perspective. How did niches emerge in the context of the existing regime? To which ongoing processes at the regime and landscape levels did the niche link up? Both questions will be addressed below.

An important claim of the multi-level perspective is that new radical technologies emerge in specific niches in the context of the existing regime and landscape. This claim was substantiated in the case study. Bicycles emerged in the 1830s in the entertainment niche, as a hobbyhorse or a 'toy' for the upper classes. At the end of the 1860s a new application domain for the bicycle emerged: bicycle racing. The selection criteria in this niche stimulated the emergence of the Ordinary, a bicycle with very large front wheels. In the 1890s, safety bicycles became linked to touring in the countryside. Steam, electric and gasoline automobiles also emerged in the entertainment niche, as fascinating technical novelties for enthusiasts in the 1880s. In the 1890s the different cars became linked to different niches. Electric vehicles emerged in the context of the electric tram regime. In this context, they were initially interpreted as a way to plug the gaps in the electric tram network, particularly as a semi-public taxi. The electric taxi niche did not survive, however. Electric vehicles also emerged in the context of the luxury carriage regime, where they were developed and used as electric luxury vehicles, for example, for tea parties, or promenading in parks and along boulevards. Internal combustion engine vehicles emerged more in the context of the bicycle, and were first used for racing and for touring. Extending the concept of the bicycle, they represented a new mobility practice, 'mobility for fun', driving for adventure, excitement and thrills. Steam automobiles were used to some extent in racing and touring. By 1903 a division of labour had emerged: electric vehicles were used as *urban* luxury vehicles; gasoline vehicles, and steamers to a lesser extent, opened up new application domains: racing and touring in the countryside. These domains were very popular, because they linked up with wider landscape developments, for example, values such as entertainment, excitement, speed, adventure and hygiene.

The other claim of the multi-level perspective was that the regime should not just be seen as a barrier for radical novelties. Established regimes may be stable, but they are not inert. In socio-technical regimes and landscapes there are ongoing processes which may provide windows of opportunity to which niche innovations can link up.

The case study did indeed show that automobiles linked up with several ongoing processes in the transport regime. First, the gasoline automobile linked up with new user preferences, which had been articulated in interaction with the bicycle (individual and flexible mobility). The automobile also linked up with the application domains of racing and touring. Second, it linked up with the trend of suburbanisation and greatly reinforced it. Suburbanisation was a development which started with the horse tram, and was strengthened by the electric tram; it came to be seen as the solution to urban social and health problems. Policy makers and urban planners embraced the automobile as a means for the wider goal of suburbanisation. Third, automobiles linked up with cultural changes in the experience of speed, which was exhilarating and exciting. This experience had already gained ground through driving carriages in parks or fast cycling. Furthermore, familiarisation with the electric tram ensured that urban residents had become more accustomed to vehicles travelling at higher speeds on the streets. This reduced the urban resistance to automobiles in the early twentieth century that had occurred with steam automobiles in the 1870s and 1880s. Fourth, the automobile could be driven on streets with smoother surfaces (for example, concrete, asphalt). The Good Roads lobby and bicycle clubs had literally prepared the way for cars. Fifth, the automobile linked up with the trend to define the function of the street as a transport artery. Moreover, it further reinforced this trend, leading to a new urban transport system in the 1910s. Sixth, it linked up with an increase in the power of public authorities in the late nineteenth century. This also included authority over streets and street improvements, which had traditionally been in the hands of local residents. Public authorities were thus in a position to promote the use of the automobile by making further road improvements.

In the transition, external landscape developments also played a role. In fact, the transition took place in a society in flux, and automobiles linked up with several landscape developments. Important social developments were urbanisation and immigration, which stimulated the growth of American cities. This rapid growth led to the spread of slums, and social and health problems. Public health was seen as an important issue. Horse-drawn transport came under pressure because it linked up negatively with this issue. Bicycles and automobiles, on the other hand, were seen as positive ways to experience nature and inhale fresh air. A cultural development was the promotion of suburban living as a cultural ideal. Middle-class American families perceived suburban homes as a haven from the commotion in the city. Urban social reformers saw suburbanisation as an answer to the social problems associated with slums. They saw electric trams as a means to this higher end. Automobiles were also linked to this goal in the early twentieth century. Another cultural development was the emergence of a new popular culture, focused on entertainment, excitement, adventure and outdoor activities. In this context, car racing and touring in the countryside became very popular. A wider political development was the expansion of public administration, taking on greater responsibility for regulating society. To demonstrate their power, cities encouraged the creation of beautiful buildings and grand concrete boulevards (American City Beautiful movement). These boulevards stimulated cycling and functioned as demonstration grounds for early fragile automobiles. Another political development was the Good Governance movement. Debates about corruption provided a suitable context for the rise of city planners, public health specialists and highway engineers. City planners and highway engineers acquired powerful positions in policy making, allowing them to promote road building.

In sum, I conclude that the multi-level perspective has a good match with the case study, and therefore is robust.

Co-evolution of Multiple Technologies

The system innovation in urban transport took place because of interactions between multiple technologies. One pattern, already mentioned in Chapter 3, was the importance of incremental innovations alongside radical innovations. Improvements in the piston engine required many incremental innovations, for example, improved carburettors, size and configuration of cylinders, placement of valves and camshafts and electric starters. Other parts of the car also embodied a wide range of technical innovations, for example, quick-drying paints, better rubber for tyres, closed bodies, electric lighting, all-steel cars and so on. In addition, the mass-production system, pioneered by Ford, was also continuously improved through all kinds of incremental innovations, thus driving down prices.

A second pattern was cross-sectoral technology flows. Technical innovations were not limited to the transport sector. The car industry posed technical questions to disciplines such as metallurgy, electrical engineering and organic chemicals. Also, use of the car called for innovations in the petroleum industry and construction and public works (roads, bridges, tunnels).

A third pattern was the importance of material innovations: quick-drying paints were important for the mass production of the Model T; float glass was a new type of high-quality glass, used in cars in the 1920s.

A fourth pattern is the 'borrowing' of technical elements from sometimes competing technical trajectories. In 1911, gasoline vehicles borrowed the electric starter from the trajectory of electric vehicles. The improved gasoline car was able to invade the urban application domain. Another example is that early electric vehicles borrowed elements from several domains: (a) batteries and battery chargers came from stationary electric applications; (b) electric motors came from the electric tram; and (c) the mechanical 'controller', a device which allowed easy gearing, was also borrowed from electric trams. This pattern means that there can be cross-fertilisation between technical trajectories.

A fifth pattern is that new technologies may act as catalyst, opening up a given regime and facilitating further change. In the literature review in Chapter 2 (Section 2.3) I quoted Sørensen (2002), who proposed the concept of a catalyst. In the transition to automobiles, the bicycle and the electric tram functioned as catalysts, in the sense that their introduction led to wider changes in the urban transport system. In retrospect, we can say that these changes prepared the way for the automobile.

The diffusion of the bicycle in the 1890s was accompanied by the following transformations in the transport system. First, the bicycle led to an articulation of the user preference for individual and flexible transport. It also opened new application domains, which later became linked to early automobiles: touring in the countryside and racing. Second, the bicycle gave rise to bicycle clubs, which put pressure on public authorities to improve the road infrastructure (the Good Roads lobby). More asphalt roads were created in the 1890s to facilitate cycling and street sweeping. This literally prepared the way for automobiles. Third, early automobiles made use of technical elements, which had been developed for bicycles, for example, steel-tube frames, the chain drive and differential gearing and air tyres. Fourth, the bicycle industry was also important in the development of techniques of quantity production utilising special machine tools, for example, sheet metal stamping and electric resistance welding. These new process technologies later became essential elements in the mass production of motor vehicles.

The electric tram was also involved in wider transformation processes in the 1890s. First, the electric tram contributed greatly to suburbanisation, a trend to which the automobile later linked up. Second, interurban trams stimulated tourism for the lower classes, allowing them to make day trips to the countryside at the weekend. This contributed to the development of tourism as a popular recreational practice. Tourism and touring became an important application domain for the automobile in the early twentieth century. Third, the trolley stimulated a change in perception of the function of streets. As more social groups lived in suburban areas, streets were increasingly defined in terms of transport. The automobile later linked up with this trend, and reinforced it. Fourth, the trolley contributed to a cultural change in the experience of speed. By 1900, speeds of 15–20 miles per hour no longer seemed quite so dangerous to many people. Because people had become accustomed to higher speeds, there was less urban protest against automobiles in the early twentieth century.

In sum, the societal embedding of the bicycle and electric tram in the 1890s contributed to wider changes in the socio-technical regime. These changes preceded the diffusion of the automobile and in many respects prepared the way for its diffusion. Hence, the automobile fell on fertile soil. From this wider and longer-term perspective it is clear that it was a last step in a wider and longer transformation process (see Figure 5.8). This means that historical descriptions, which focus on the emergence and diffusion of automobiles (as indicated with the dashed circle), constitute only a limited cross-section of the wider transformation process.

Diffusion as a Trajectory of Niche Accumulation

Steam, electric and gasoline automobiles diffused because they were used in subsequent application domains. Early experiments with electric vehicles (EVs) in the 1840s failed, because batteries were too heavy. The improved batteries in the 1880s led to a new wave of technical novelties. These were used as taxi and society vehicles (in parks, for tea parties). Attempts to enter the niche of racing and touring failed. In the 1830s, attempts to use steam engines in buses in the London area resulted in little success, because steam engines were heavy and cumbersome. In the 1860s and 1870s, steam

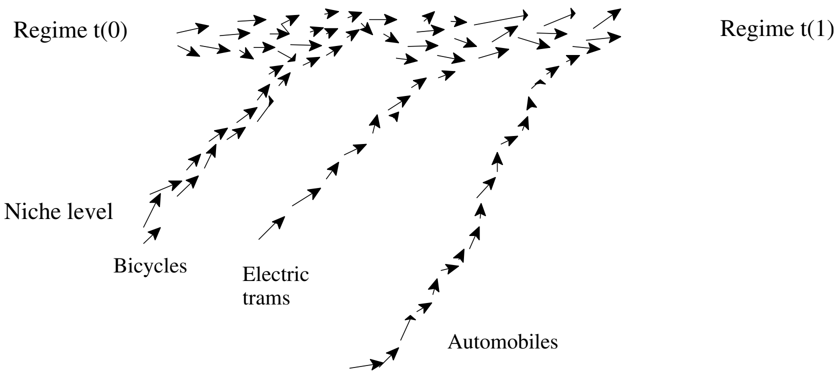


Figure 5.8 Automobiles as last step in a longer transformation process

engines became smaller and stronger (partly because of the use of steel). In America, small and light steam automobiles were built, but there was resistance to using them on roads, because people objected to the high speeds. In the late 1880s the flash boiler made steam automobiles easier to operate. In the 1890s steam automobiles were sold and used, because of changes in the perception of the function of streets and because people had become accustomed to faster vehicles. Steam automobiles were used to some extent in races and touring.

The gas engine was developed for stationary applications (for example, in machine shops). In the 1880s it was applied to tricycles (Benz) and coaches (Daimler). In the late 1890s, gasoline automobiles became popular in the niches of racing and touring, that is, for fun and entertainment. Doctors, salesmen and rich farmers were the first to use gasoline automobiles for business functions. The saturation of the luxury niche led to the sales crisis of 1907, which stimulated the emergence of a new design trajectory: cheap and strong cars. Ford's Model T (1908) became the dominant design in this trajectory. It was used as a gasoline taxi in cities and increasingly by farmers. In the 1910s, the countryside provided a major market niche for gasoline cars. As the electric starter (1911) made the gasoline car easier to operate, it invaded the urban niche, which had been the domain of the electric vehicle. In the urban niche, electric vehicles had even enjoyed a 'golden age' (1907–11). With the introduction of the electric starter, the gasoline vehicle pushed the electric vehicle from its last niche. Gasoline cars were increasingly used for commuting by middle-class suburbanites. In the 1920s the car became an all-purpose road cruiser, which began to challenge the electric tram as the dominant urban transport

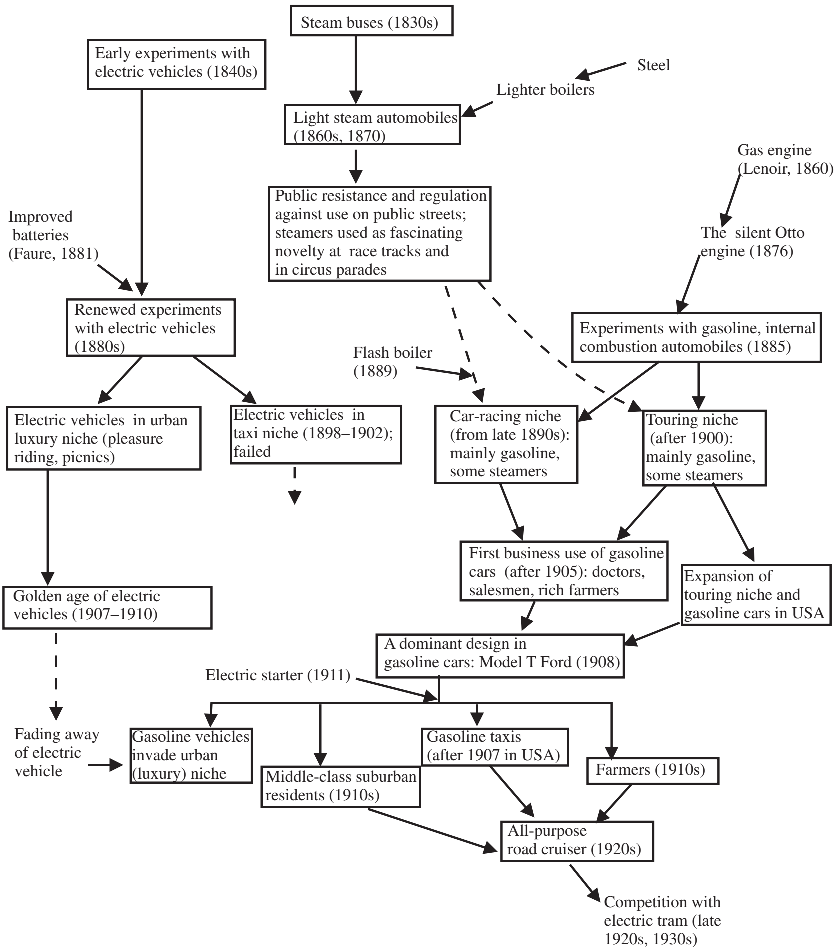


Figure 5.9 Trajectory of niche accumulation for automobiles

technology. This trajectory of niche accumulation is schematically represented in Figure 5.9.

Phases in the Innovation Journey of Novelty

In Chapter 3, I distinguished four phases in system innovations. In the automobile transition these four phases did indeed occur, and therefore I conclude that they are corroborated.

First phase: emergence of novelty in the context of existing regime and landscape

Electric vehicles were initially interpreted in terms of the electric tram regime (1880s). As an electric taxi it was supposed to plug the gaps in the electric tram network. Its functionality was also interpreted in the context of the luxury coach regime, namely as a society vehicle for promenading in parks, and for tea parties. The functionality of the gasoline vehicle was interpreted in the context of the bicycle regime, namely for racing and touring (entertainment). In a technical sense the first automobiles were close to existing forms, made by adding engines or motors to existing coaches or tricycles.

Second phase: technical specialisation in market niches; exploration of new functionalities

During the 1890s, gasoline cars developed a characteristic form of their own: they were given a nose on the front, and became lower and wider (stretch). The form of EVs remained very similar to the existing coaches. While the early automobile producers were bicycle makers or coachbuilders who undertook automobile manufacturing as an additional activity, a dedicated automobile community emerged in the late 1890s. As the market niches of racing and touring expanded, gasoline cars helped to spread a new mobility practice: touring for fun (entertainment, excitement) in which drivers enjoyed the new functionalities of individual and flexible mobility at high speeds.

Third phase: breakthrough, wide diffusion and competition with the established regime

In the first years of the twentieth century, a new design trajectory emerged: cheap and strong cars. After the sales crisis of 1907, this trajectory was embraced by the automobile community. Ford's activities resulted in the emergence of a new dominant design, the Model T (1908).

Economic perspectives emphasise that the further diffusion of the gasoline automobile was driven by price/performance improvements. The price of Ford's Model T decreased from $850 in 1908 to $360 in 1916, which helped to make the car affordable for more consumer groups (first the middle classes, later the upper working classes). The performance of the car was also improved with many incremental innovations, for example, electric starter (1911), lighting systems, more powerful engines and an enclosed body (1910s) and door locks (1912).

Socio-technical perspectives emphasise that diffusion took place because an increasing number of heterogeneous elements were linked together. In the 1910s and 1920s a new urban transport system was created with the development of new roads, pavements, traffic and behavioural rules, traffic signals and lights, and public organisations (for example, highways departments).

Sociological perspectives emphasise the social struggles that went on and how the automobile was a means to a higher end or a pawn in a wider game. One example is that progressive policy makers embraced the car not for its own sake, but because they saw it as a tool to stimulate their goal of suburbanisation. Another example is that highway engineers took advantage of the distrust of policy makers to portray themselves as rational and objective experts. In this way they acquired powerful positions in public bodies, allowing them to promote road building in the 1920s and 1930s. A third example is that local authorities were engaged in a long-standing struggle with tram companies, whose public image was one of greedy monopolists. In the context of this struggle they chose to sponsor the car (through favourable regulation and road investments), at the expense of the electric tram. Likewise they embraced the motorbus, which put more pressure on tram companies in the 1930s.

Fourth phase: gradual replacement of established regime, transformations and wider impacts

The declining price of cars in the 1910s and 1920s, as well as the creation of a new socio-technical regime, stimulated the diffusion of the automobile in rural areas and among the urban middle class. In the 1920s the automobile began to compete with the electric tram, stimulated by political choices and (financial) regulations. Urban planners and highway engineers helped to construct a car-based transportation system. Existing roads were widened and new roads were created, for example, highways. Automobiles were given the right of way and pedestrians and cyclists were pushed to the side of the road or onto sidewalks.

A major impact of the car was the transformation of the city, in particular the advent of suburbanisation. The automobile also contributed to a major transformation of rural areas. Institutions such as schools, churches, shops and medical facilities disappeared from small villages.

Fit–Stretch Pattern in the Co-evolution of Form and Function

In a fit–stretch table, I have summarised the co-evolution of form and function during the automobile transition (see Table 5.5).

Table 5.5 Fit–stretch in the co-evolution of form and function in the automobile transition

Use environment \ Technical form	Fit	Stretch
Fit	Early automobiles were carriages or tricycles with power sources. Electric vehicle interpreted as: (a) 'plugging the gap' of electric tram (for example as taxi (b) substitute for luxury horse and carriage. Gasoline cars build upon bicycles: races and touring	Gasoline car developed a specific form in the context of races and touring: lower, wider, with 'nose' (the hood).
		In America touring became very popular (mobility for fun), exploiting the new functionalities of private, flexible transport
Stretch		After 1905 the automobile was used for business purposes. Model T Ford (1908) was new dominant design: cheap, robust cars for practical use. Scale effects and mass production drove the price down, stimulating the wider diffusion, which was accompanied by changes in infrastructures, traffic rules, pedestrian routines, mobility patterns

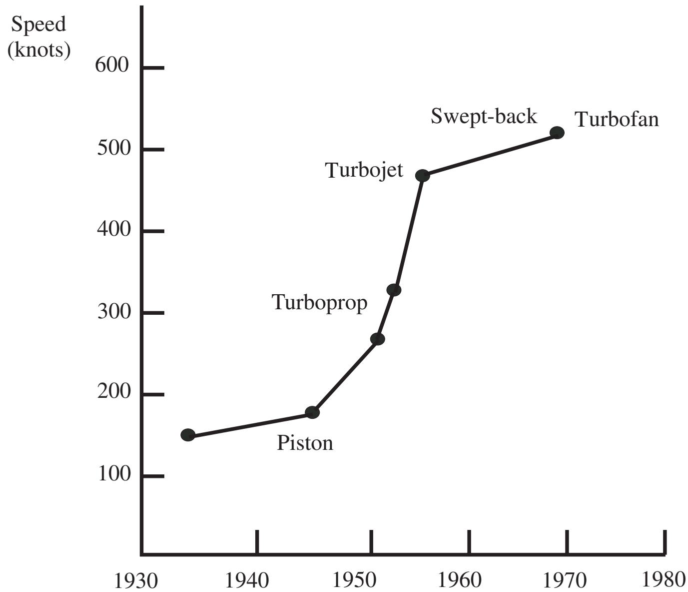
6. The transition from piston engine aircraft to jetliners in American aviation (1930–1975)

6.1 INTRODUCTION

This chapter analyses the transition from piston engine and propeller aircraft to jetliners in America from 1930 to 1975. This transition involved not only technical changes in engine components, but also changes in the aircraft as artefact, for example, wing structures, airfoil and electronics. The transition had further knock-on effects throughout the aviation system. The jet engine not only changed the form of aircraft, but also their performance characteristics. Because jet engines had more thrust than piston engines, they enabled a substantial increase in speed (see Figure 6.1). The dominant passenger aircraft in the mid-1930s, the DC-3, was capable of carrying 14–21 passengers, had a range of 1000–1500 miles and a speed of 160–190 mph (250–305 kmph; 130–157 knots). In 1958 the jet-powered Boeing 707 raised cruising speeds to 550 mph (885 kmph; 480 knots), accommodating up to 181 passengers and with a range of about 3000 miles. In 1969 the Boeing 747 raised the number of passengers to about 450–500, and had a range of about 6000 miles and a cruising speed of 600 mph (965 kmph; 520 knots).

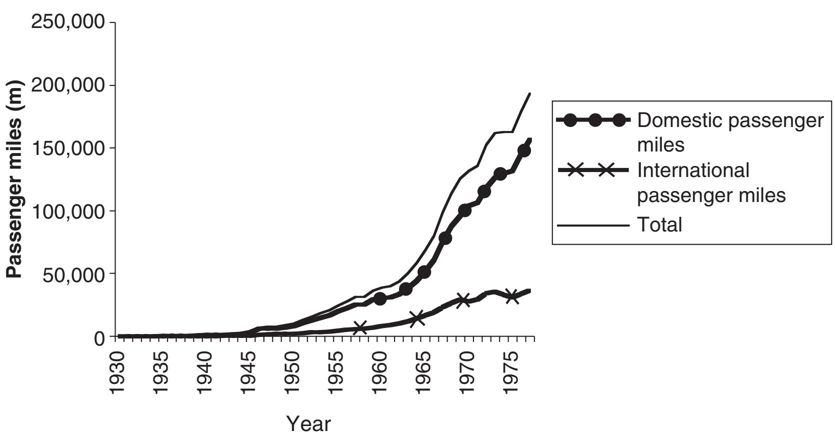
The jet engine not only allowed higher speed, but could also power larger and heavier aircraft, carrying more passengers, thus allowing economies of scale. Lower airfares and shorter travel times stimulated the expansion of passenger markets. Figure 6.2 shows that the transition from piston engine aircraft to jetliners coincided with a tremendous growth in American passenger transport. In particular, the wide-bodied Boeing 747 introduced a new functionality in flying: mass transportation. The cultural perception of flying changed from something for daredevils to a normal way of transportation.

Jet engines also affected military aviation, enabling fighter aircraft to go faster than the speed of sound. High speed is a crucial performance characteristic for fighter aircraft, and it was the main reason why the military were willing to sponsor the development of jet engines during the Second



Source: Hopps (1978: 102).

Figure 6.1 History of cruise speed in civil aviation (in knots)



Source: Data from web page of the American Air Transport Association (accessed 23 February 2004).

Figure 6.2 American civil aviation, domestic and international, 1930–1977

World War. Early jet engines had the disadvantage of a high fuel consumption, and the higher fuel costs made them unattractive for civil aircraft. But for fighter aircraft, especially interceptor fighters, high fuel use was less important than high speed. Jet engines also changed the design of bombers, enabling them to fly at high speeds and high altitudes. Together with tanker aircraft, jet bombers changed war strategies during the Cold War. The infamous B-52 heavy bomber assured mutual destruction if the Soviet Union were to attack the United States.

The increasing traffic volume, enabled by the jet engine, caused problems in the sky. The increasing speed difference and congestion at airports required the upgrading of air traffic control systems. Because jet aircraft were heavier, they required other changes in airport infrastructure, for example, lengthening and strengthening runways. A problem of societal embedding was the noise from jet engines. Civil protests put the issue of noise on the political agenda, leading to new regulations. In sum, the technological transition to jet engines was part of a wider system innovation in aviation. This chapter analyses how the transition and system innovation came about.

The jet engine was developed in Britain and Germany before the Second World War. During and after the war America and the Soviet Union also acquired turbojet knowledge and applied it to fighters and bombers. Britain pioneered the use of jet engines in civil passenger aviation, but it failed to achieve long-term success. However, civil jetliners were successfully introduced in America in 1958 with the Boeing 707. The analysis in this chapter looks in more detail at the strategic positioning and interactions between these countries. But most attention is given to America, because it succeeded in introducing civil jetliners and presided over the expansion of aviation. Both the military domain and civil aviation will be described.

To analyse the transition to jetliners and the wider system innovation, I shall adopt the conceptual perspective described in Chapters 1 and 3. First, I shall give an indication of the socio-technical system in aviation, to get an impression of the relevant elements. Figure 6.3 presents a sketch of this system, filled in for jet aircraft. Many of these elements did not exist, but had to be created and linked together. I shall analyse how the elements and the linkages interacted and co-evolved. Although engine components receive much attention in this chapter, the analysis will also address the wider aviation system.

I also present a sketch of the main social groups involved in aviation (Figure 6.4). These groups maintained (and changed) the elements of the socio-technical system. They acted in the context of a set of rules, the socio-technical regime, which involved design rules, problem agendas, search heuristics, dominant designs, landing rights, user practices and so on.

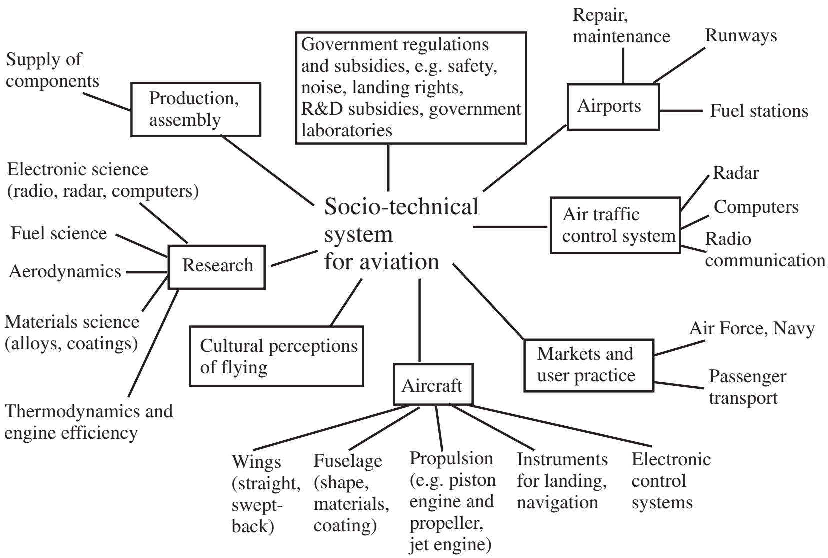


Figure 6.3 Components and linkages in socio-technical system in aviation

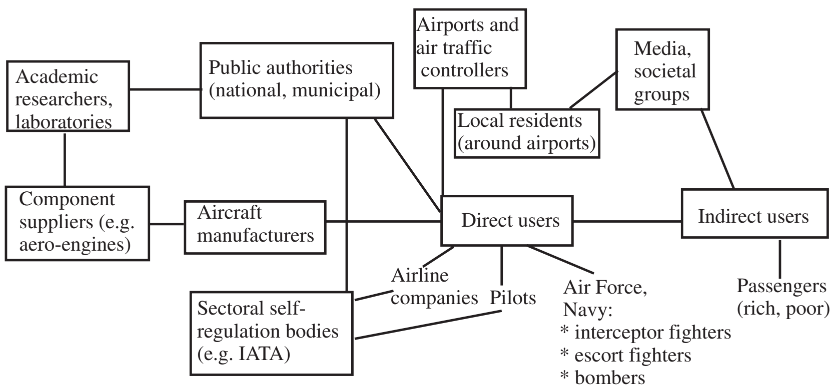
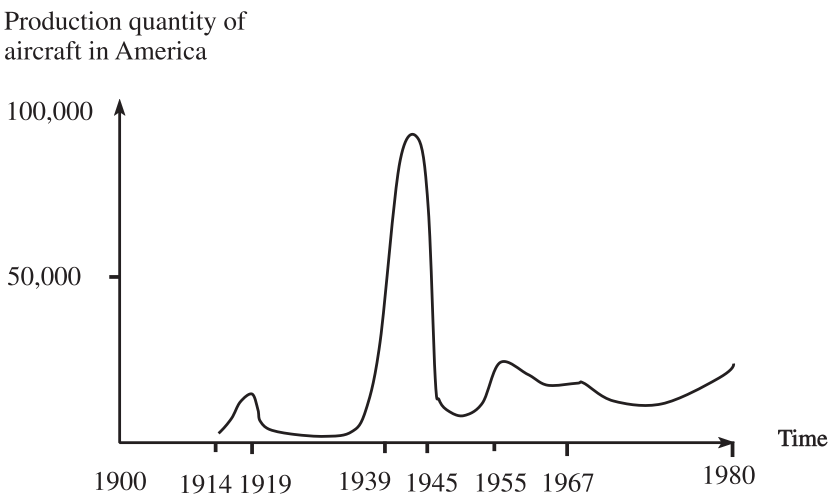


Figure 6.4 Social groups in aviation

These indications of system, social groups and regime show that the transition was complex, involving many elements, actors and rules. The question is how interactions between them resulted in a system innovation. To structure the analysis I shall use the multi-level perspective, described in Chapter 3. This means that the analysis pays attention not only to the level of regime and system, but also to the niche level, where novelties emerge.

We shall see that the gas turbine emerged in the context of electricity generation, where the focus was on rotational power rather than thrust. In the 1930s, several academic pioneers abandoned this idea and calculated that the thrust of gas turbines could be used to power aircraft. Because these pioneers challenged accepted ways of thinking, they found it hard to attract support for their ideas. It was not until the approach of the Second World War, that military planners became more interested. With development subsidies they created a niche for work on jet engines for a particular kind of application: interceptor fighters. The performance of these fighters during the war surprised the aviation community. After the war, jet engines gradually entered other market niches. The third level of the conceptual perspective is the wider, external landscape. In this case study the main landscape developments are wars: the Second World War, the Korean War (1950–53) and the Cold War. These wars strongly influenced the production of aircraft (see Figure 6.5).

The hypothesis is that the interaction between these three levels explains the timing and course of the system innovation. This hypothesis will be tested in the concluding section of the chapter. The transition to jet engines has previously been described by others, for example, by Constant (1980). But these other descriptions follow a point-source storyline, focusing on the emergence and diffusion of the jet engine. My analysis is different in the sense that I pay a good deal of attention to the



Source: Todd and Simpson (1986).

Figure 6.5 American aircraft production existing piston engine-propeller regime. I shall show how turbojets first emerged in this existing context and gradually broke free from it. I shall tell the story in a different way, showing the added value of the multi-level perspective.

Another aim of the chapter is to test the patterns described in Chapter 3: co-evolution between multiple technologies, diffusion as a trajectory of niche accumulation, phases, and fit–stretch pattern in the co-evolution of form and function. These patterns will be explicitly addressed in the concluding section.

The empirical description is organised along the four phases in system innovations. Together with an additional section on the pre-history, the story is told in five sections, addressing different periods. The first period (1920–36) is characterised by the creation of a technical regime in aviation, resulting in the piston engine DC-3. In the context of this emerging regime, academic pioneers developed their first ideas about jet engines (Section 6.2). The second period (1936–40) is characterised by the creation of a socio-technical civil aviation regime around the DC-3. At the same time the approach of war led to rearmament programmes and more attention was directed towards the development of the jet engine (Section 6.3). In the third period (1940–45), jet engines moved from idea to prototype to actual production and use in interceptor fighters. This occurred in the context of the Second World War (Section 6.4). In the fourth period (1945–60), jet engines were further improved in the military domain, for use in fighters and bombers. In the civilian domain there was much uncertainty about the engine of the future. While jet engines offered more thrust, but higher fuel consumption, improved piston engines promised higher fuel efficiency; the turboprop was a third engine, which seemed to promise the best of two worlds. The uncertainty created space for strategic manoeuvring. Britain took the lead in pioneering civil jetliners, while America adopted a wait-and-see attitude. The British jet-powered Comet failed, but nevertheless triggered America into action. Strategic games between Boeing and Douglas Aircraft ended in the victory of the Boeing 707, ushering in the jet age (Section 6.5). In the fifth period (1960–75), piston engines were gradually replaced by increasingly fuel-efficient jet engines. The diffusion of jetliners in civil aviation triggered a range of adaptations in the aviation system (Section 6.6). The chapter ends with an analysis and conclusions (Section 6.7). For each section the following dynamics are described:

- Relevant external landscape developments.
- Main developments in the socio-technical regime, for example, technological developments, market dynamics, policy actions, problems

and tensions. How were these regime processes influenced by landscape developments?

- Emergence and development of novelties in niches, and the actors involved, the problems they encountered, and their attempts to overcome them. How were niche dynamics related to ongoing processes at the regime and landscape levels?

6.2 CREATION OF AN ALL-METAL, PISTON ENGINE AIRCRAFT REGIME (1920–1936)

Landscape Developments

An important landscape development of the 1920s was economic and societal optimism. The Roaring Twenties were characterised by an explosion of image- and sound-making technologies (radio, phonograph, film). It was an era of experimentation with new forms of entertainment (for example, dance halls, movie theatres). During the 1920s, spectator sports enjoyed an unprecedented growth in popularity. As radio, newspapers, magazines and newsreels exhaustively documented their exploits, athletes and sportsmen and women took their place alongside movie stars. The sense of adventure was accompanied by a desire for progress and modernity. There was a feeling of entering a new age, represented by modernity. Another aspect of the cultural climate was technological utopianism (Segal, 1986), the belief in technology as the means of achieving a perfect society in the near future. Aircraft and aviation linked up well with this cultural climate. Air races in the 1920s were very popular and attracted many spectators. This was accompanied by visions in the 1920s and 1930s about the coming air age, in which aircraft would change society for the better. One example was the vision of mass personal flying. Prophets considered the aeroplane to be the 'horseless carriage of the next generation' and wrote enthusiastically about the prospect of an 'aeroplane in every garage'. In sum, there was widespread cultural enthusiasm about aviation, the so-called 'winged gospel' (Corn, 1983). This enthusiasm provided a 'cultural space' for government support.

Another landscape development was the stock market crash of 1929 and the subsequent Great Depression, leading to mass unemployment and economic slump. To combat the excesses of poverty and unemployment, governments around the world increased public spending on labour-intensive projects. In America, the New Deal (1933–35) provided the context for reconstruction projects. The construction of airports was one of those projects, thus stimulating the infrastructural basis for aviation.

Creation of a Technical Regime in Aviation

In the 1920s and early 1930s aviation was still a small and insignificant transport mode in terms of passenger miles. Nevertheless, it was supported by cultural enthusiasm and high public visibility. Although there was not yet an integrated aviation system, there was much innovation in many elements, for example, engines, airframes, airports, navigation instruments, aerodynamic science and markets. This section describes how these elements gradually linked up. An important step was the DC-3 (1936), which formed a technical aircraft regime. Other elements could be organised around this technical core.

In the 1920s there were several application domains for aircraft. One market niche was commercial passenger transportation. The first airline service for passenger transportation started in 1919, between London and Paris. But commercial aviation remained small scale, and was not profitable. Public authorities stepped in to subsidise the airlines for a variety of reasons (imperial policy, military, political prestige). European governments preferred direct subsidies to airlines, while American governments used indirect airmail subsidies. In many European countries the subsidies amounted to between 50 and 80 per cent of the income of European airline companies (Miller and Sawers, 1968: 15). A second market niche, especially in America, was mail transport. With the Contract Air Mail Act of 1925, airline companies were paid generously on a per-pound basis (Van der Linden, 1995: 247). Airmail subsidies provided substantial support for commercial aviation. For the 1931–39 period, airmail constituted a major portion of American airline income: $134,276,000 in airmail receipts versus $144,476,000 in passenger revenues (Constant, 1980: 168). Because private mail carriers were interested in planes with low operating costs, aircraft manufacturers built specialised mail carriers, for example, the Boeing Monomail. Another regulation, the Air Commerce Act (1926), was a further stimulus for aviation, mandating the federal government to be a promoter and regulator of aviation, that is, to license pilots, inspect and register aircraft, control air space and establish civil airways. Entertainment and air races formed a third application domain. The 1920s were the great age of aeronautical competition and aviation shows, including acrobatics and air races (ibid.: 169). The late 1920s were also the era of long-distance flights. The solo flight by Charles Lindbergh across the Atlantic Ocean in 1927 boosted cultural enthusiasm for flying in the United States. The air shows and races did much to foster and spread air-mindedness, and pushed forward technical developments, for example, the shift from bi- to monoplanes. The niche also turned the attention of scientists to the problems of high-speed and high-altitude flying. A fourth application domain was

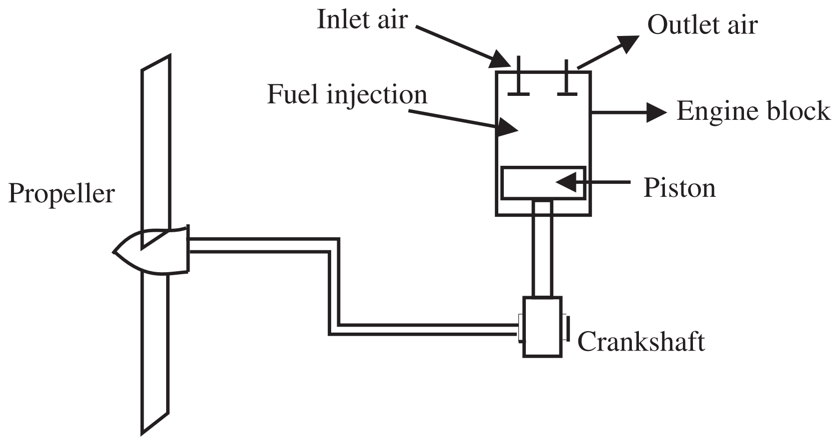


Figure 6.6. Schematic representation of piston engine and propeller

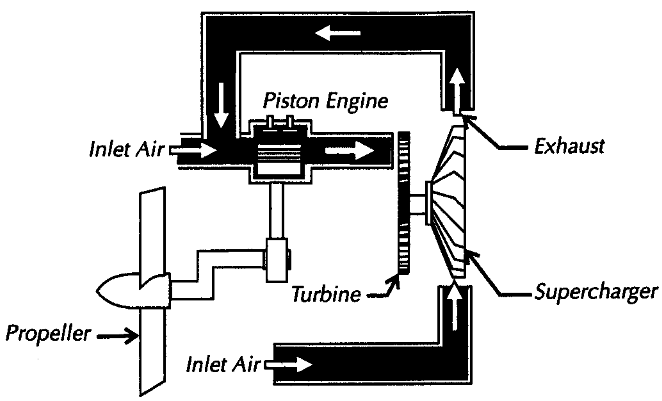
military. Although the niche was small, the Army spent money on innovative work, particularly regarding the shift from wooden to metal aircraft structures to protect pilots from gunfire.

Aircraft in this period were propelled with piston engines and propellers. The reciprocal movement of the piston engine was translated in a rotary motion with a crankshaft, which drives the propeller (see Figure 6.6). In the 1920s and 1930s great improvements were made in piston engines, for example, in gross power output, power per pound, specific fuel consumption, altitude and speed. Performance improvements stemmed from an accumulation of small innovations, for example, cylinder blocks, better pistons, internally cooled valves, fuel innovations (ibid.). Development of better engines also entailed development of better accessories, for example, fuel pumps, magnetos, electrical systems, sealants, lubricants and controls.

Despite tremendous progress, piston engine propeller aircraft encountered two problems. One was the high-altitude problem. Because air gets thinner at greater heights, propellers have less grip and engines suffer a lack of oxygen, reducing the efficiency of the combustion process. The second was the high-speed problem. At high velocities the speed of propeller tips approached the sound barrier, leading to the formation of a ‘compressibility burble’, which substantially increased drag and decreased speed.

There was a strong belief in the aviation community that both problems could be solved with incremental innovations. One response to the altitude problem was the development of variable-pitch propellers, making it possible to change the propeller pitch in flight, adjusting it for differing speeds, altitudes and loads. A response to the engine-altitude problem was the adoption of superchargers and turbo-superchargers. By compressing the less dense air at high altitude, piston engine power could be maintained. Supercharging refers to the pre-compression of air ingested by a reciprocating piston engine. Turbo-supercharging refers to the use of a gas turbine in the exhaust stream of a piston engine, thus using the otherwise wasted energy (see Figure 6.7). Gas turbines thus formed an auxiliary arrangement in the piston engine. The altitude problem created a niche for the application of gas turbines within existing aircraft. The experience with gas turbines in turbo-superchargers (dealing with exceptionally high rotational speeds, up to 40,000 revolutions per minute (rpm), at high temperatures, up to 600°C), would later be useful for developing turbojets (ibid.: 125).

Progress in aircraft stemmed not only from improvements in piston engines, but also from smoother airframes, which made use of insights from aerodynamic science. Aerodynamic science made major conceptual advances. Robert Thomas Jones found in 1927 that the main ingredient in the drag of streamlined bodies was 'skin friction'. This highlighted how much power was wasted by existing aircraft (especially biplanes). In 1930, Theodore von Karman proposed the concept of 'turbulent drag' to understand the difference between streamlined and turbulent flow. These advances laid the foundation for the radical redesign of airframes. Aerodynamics also made advances with regard to supersonic flight. Experiments with 'compressibility burble' indicated that there was a limit to the speed at which propellers could be used. Insights into supersonic flow behaviour emphasised the importance of aircraft streamlining.



Source: Heppenheimer (1995: 80); courtesy of John Wiley & Sons.

Figure 6.7 Turbo-supercharger

Much could be done to better streamline aircraft, most of which, in the 1920s, were still biplanes, consisting of wooden or metal frames covered with linen. The application domains of racing and the military, with their emphasis on speed and range, provided opportunity for experiments with monoplane designs, which reduced drag and heightened the aerodynamic efficiency. Two other innovations also revolutionised aircraft design. The first was a change in the primary construction material from wood to duralumin (Schatzberg, 1994), an aluminium alloy, which combined strength and lightness (Rae, 1968a: 13). The second was the introduction of the monocoque airframe, that is, a structure, which both carries the loads and shapes the aircraft. The aluminium skin itself was made to carry a significant portion of the structural load.

Airports in the 1920s were a mixed bag, ranging from grass fields to complex infrastructures. These airports were created by municipalities or private business. The US government had no influence, because the Air Commerce Act of 1926 did not authorise their involvement in building or operating airports. Franklin Roosevelt's New Deal signalled a major change in federal policy towards airports, leading to an expansion of the network of airports. Between 1932 and 1938 about 75 per cent of funding on airport development came from the New Deal (Heppenheimer, 1995: 119).

Another aspect of the emerging aviation system was navigation technology, helping pilots to determine their course and find their way between airports. In the late 1920s some basic aircraft instruments were developed, for example, the artificial horizon, the directional (heading) gyro and altimeters. Cockpit instruments were further improved in the early 1930s, with better altimeters, airspeed indicators, rate of climb indicators, and the 'artificial horizon', which showed pilots the altitude of the aircraft relative to the ground. To facilitate night flying, a lighting system was created by placing lines of large electric beacons at 15-mile intervals, stretching from one airport to another. By 1933 some 18,000 miles had been illuminated (Douglas, 1995: 65). These new instruments made flying safer and also allowed operation in bad weather, thus improving flying on schedule, something which passengers greatly appreciated.

The wide range of innovations in the aviation system formed the basis for the rapid growth of US commercial aviation in the 1930s. Aircraft with better engines, airframes and navigation instruments were safer, faster and more punctual, reducing passengers' fears. Passenger transport was also stimulated indirectly with mail subsidies, which ensured profitability for the airline companies. Furthermore, the long distances on the American continent created a market for innovative planes with an ever-longer range. These distances could be flown within the same country, preventing hassles over sovereign air space, landing rights and passport formalities (Miller and Sawers, 1968: 18). The main users of passenger services, businessmen and politicians, made up 85 per cent of all passenger travel miles in 1933 (Douglas, 1995: 73). The titans of business were not immune to the dramatic imagery of aviation. To fly was to identify with the modern.

Profitable mail subsidies and growing passenger markets provided airline companies with money to demand better aircraft. New aircraft were developed in the early 1930s with novel design features, for example, all-metal, mono-wing planes, using stressed-skin structures, retractable landing gear, Fowler air flaps and leading-edge slots. The accumulation of these design changes resulted in much stronger and more aerodynamic aircraft, the so-called 'airframe revolution' (Rae, 1968a). Two notable new planes were the Boeing 247 and the Douglas Commercial Aircraft (DC-1, DC-2 and DC-3). The DC-3 eventually emerged as winner from the competitive game, and formed a dominant design for the aviation industry. The DC-3 was introduced in 1936, could seat 21 passengers, fly at 160–190 mph (250–305 kmph) and had a range of 1000–1500 miles. The high fuel efficiency of the DC-3 plane gave it good economic performance. By the outbreak of the Second World War, 80 per cent of US aircraft were DC-3s (Tushman and Anderson, 1986: 453).

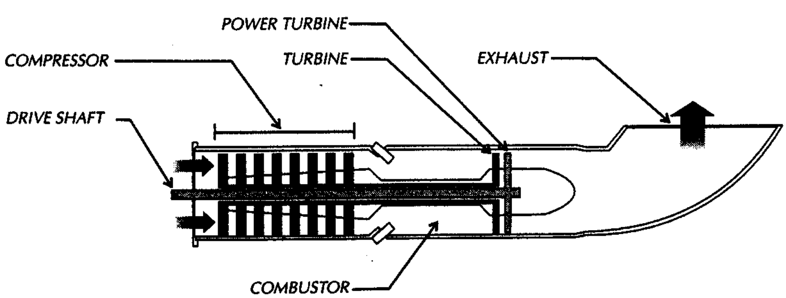
Alternative Technologies on the Niche Level

The high-altitude and high-speed problems in piston engine propeller aircraft triggered innovative work on alternative technologies. A relatively incremental innovation was to use diesel piston engines, which were safer, because they had fewer parts to go wrong. Diesel engines were also readily adaptable to high degrees of supercharging, which made them potentially well suited to high-altitude operation. Diesel engines offered good specific fuel consumption, and diesel fuel was relatively cheap (Constant, 1980: 134). During the early 1930s, diesel engines were used on a moderate scale on Junkers aircraft belonging to Deutsche Lufthansa and the Luftwaffe. The great disadvantage of diesel engines was that they were heavy because of their higher cylinder pressure. But they also failed because of improvements in gasoline aero-engines.

A more radical option was the turboprop engine, in which a gas turbine drove the propeller. The turboprop was thoroughly investigated in the 1920s. The British Royal Aircraft Establishment (RAE) commissioned an expert, Alan Arnold Griffith, to write a report about the engine in 1929. A special engine subcommittee of the Aeronautical Research Council (ARC) approved construction of a 14-stage test unit in 1930, but because of the economic depression the test was never conducted. The turboprop idea remained shelved until 1936. The aero-engine community felt that the option of gas turbine propulsion had been explored and decisively rejected. However, even while negative opinion hardened, science was breathing new life into the gas turbine.

The gas turbine, an important component of the turbojet engine, was initially developed in the context of electricity generation in the early twentieth century. In a gas turbine, inflowing air is first compressed by a compressor. Then fuel is injected in the airflow of the combustion chamber, which creates heat and expands the air. As the air flows out of the combustion chamber, it spins the blades of a turbine, which is connected to the compressor. The remaining part of the energy is used to spin the blades of a power turbine, which is connected to a drive shaft, which rotates in a magnetic field, and generates electricity (see Figure 6.8). Thus, in electricity production the focus was on the rotating energy of the drive shaft, not on the energy of exhaust gases.

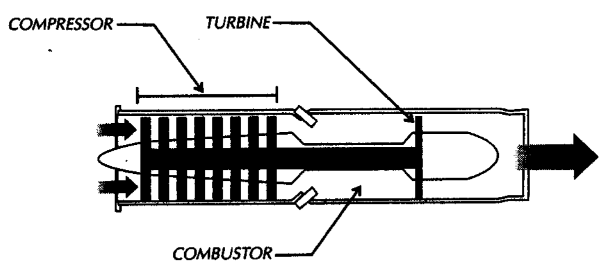
This focus on rotating shaft power also inspired the engineers who developed the turboprop: they used the rotating shaft to drive propellers. In the early 1930s, some academic pioneers shifted their focus in the study of the gas turbine. Instead of looking at rotational shaft power they began focusing on the thrust of exhaust gases. Ideas about turbojets, a radically new kind of propulsion, emerged independently from the work of three men: Frank Whittle in Britain, and Hans von Ohain and Herbert Wagner in Germany. The protagonists of the turbojet revolution were all outsiders to the normal aero-engine community, young and with an academic background. The airframe revolution in existing aircraft motivated these scientists to ponder the high-speed problem. Assuming that improvements in airframes would continue, they reasoned that near-sonic speeds would become possible in the future. They further reasoned that it would become



Source: Heppenheimer (1995: 78); courtesy of John Wiley & Sons.

Figure 6.8 Gas turbine for electricity production difficult to realise these speeds, because the compressibility burble around the propeller would increase drag. They concluded that propeller propulsion had inherent speed limits. This 'presumptive anomaly' (Constant, 1980) encouraged them to search for different propulsion systems. They eventually arrived at the turbojet engine, using not the rotational power of the gas turbine, but the thrust of the exhaust gases. Jet engines basically consist of three parts: compressor, combustion chamber and turbine. The compressor sucks in air, and increases its pressure. The compressed air enters the combustion chamber where fuel is injected and ignited. The air expands and flows out of the combustion chamber at high speed. The high-speed airflow spins the blades of the turbine, which drive the compressor. When the hot exhaust gases force themselves out of the back, they create a strong forward thrust (see Figure 6.9). According to calculations by the pioneers, the turbojet engine could be propulsively efficient at high speeds. The greater fuel consumption of the turbojet might be compensated by its lighter weight compared to a piston engine and propeller.

The turbojet pioneers had difficulty attracting the interest of (and money from) the aviation community. Whittle spoke with the Air Ministry in 1930, but was rejected. He then approached a number of firms, which also showed no interest. In 1935, Whittle and two friends formed a small company, Power Jets, Ltd, to develop the engine themselves (Constant, 1980). The commercial application envisaged by Power Jets was a small, fast, high-altitude trans-Atlantic mailplane (ibid.). At Gottingen University, Hans von Ohain constructed a small demonstration engine in 1934 at his own expense. But the combustion chambers did not function properly and the engine never really worked. At the Technical University of Berlin, Herbert Wagner and his assistant Max Adolf Muller concluded in 1937 that, without a propeller, the gas turbine could result in a light engine, with acceptable efficiency at



Source: Heppenheimer (1995: 78); courtesy of John Wiley & Sons.

Figure 6.9 Turbojet engine

high speeds and altitudes. They sought external support to develop their idea. Thus, by the mid-1930s there were several ideas for developing a turbojet. However, because there was little interest from the established aeronautic community, the jet engine was developed as a 'hidden novelty', which had to wait for better times.

6.3 CREATION OF A SOCIO-TECHNICAL AVIATION REGIME (1936–1940)

Landscape Developments

The most important change at the landscape level was the approach of war. When Hitler came to power in 1933 he accelerated the build-up of panzer divisions and the air force. By late 1935, Great Britain and France had become alarmed by Germany's political and military moves, and started their own rearmament programmes. In aviation their attention was directed towards better fighters, aero-engines and radar. While European countries were preparing for war, America was in a mood of complacent isolationism. War preparations accelerated technical developments in military aviation and made funds available for new kinds of engines, including the turbojet engine.

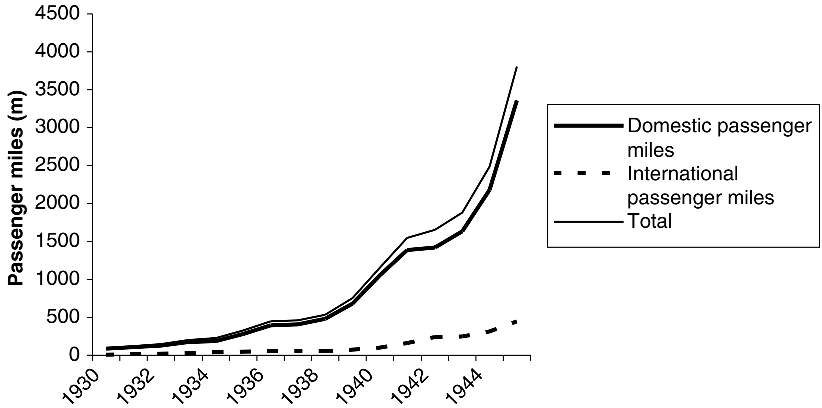
Dynamics in the Regime of Civil Aviation (1936–1940)

In civil aviation the diffusion of the DC-3 triggered a range of further innovations in passenger markets, runways, navigation technologies and symbolic perceptions. The alignment of these innovations resulted in a socio-technical aviation regime.

The DC-3 transformed air travel into a practical and reliable means of long-distance transport. American commercial aviation continued to grow rapidly (Figure 6.10), but still trailed far behind the number of train passengers. In 1938, 1,176,858 travellers used aircraft compared to 15,539,847 who used the train (7.6 per cent) (Bilstein, 1995: 97).

The American network of passenger services was expanded, offering cross-continental services, for which the DC-3 needed three or four refuelling stops. These flights were not very comfortable, and passengers complained about noise, vibration, turbulence and airsickness. Both passengers and airline companies demanded larger aircraft, which could fly above the weather to reduce airsickness. They also wanted greater speed and longer range.

In response to these user preferences, aircraft manufacturers designed new aircraft, stretching their existing models. The cumulative effect was the



Source: Data from web page of the American Air Transport Association (accessed 23 February 2004).

Figure 6.10 American civil aviation

rise of second-generation passenger aircraft with four engines in the late 1930s, Douglas's DC-4, Boeing's Stratoliner and Lockheed's Constellation. The development of these new designs was stalled by the outbreak of the war. To allow the four-engine aircraft to fly higher through thinner air a new technology was developed: pressurised cabins. But flying at high altitudes was limited by the use of piston engines and propellers, because of the altitude problem. Even with supercharging, aircraft could not fly sufficiently high above the weather. There was a conflict between the type of motor power and flying high to increase passenger comfort.

The airframe revolution, the DC-3 and the four-engine aircraft triggered changes in the production process. As aircraft became more complex, specialisation occurred in production. Specialised firms produced components (for example, fuselage, wings, aero-engines, propellers, accessories) and airframe manufacturers assembled them. The aviation industry also consolidated. From 1934 to 1940, four large airframe producers comprised the bulk of the American civilian aircraft industry (Mowery and Rosenberg, 1982: 166). Boeing Aircraft Company was a relative weakling in the commercial aviation market, but produced many bombers for the Army. Douglas Aircraft Company was the dominant civilian aircraft manufacturer because of the tremendous success of the DC-3. Lockheed established itself as a moderate player in the civilian market, but was not a major factor in the military market until 1938. McDonnell Aircraft Company was a small manufacturer, but expanded in 1939 as military demands increased. In 1967 it merged with Douglas Aircraft, forming McDonnell Douglas Corporation. The two dominant aero-engine manufacturers were Pratt & Whitney, and Curtiss-Wright, with Rolls-Royce (Britain) as a third, non-US world player.

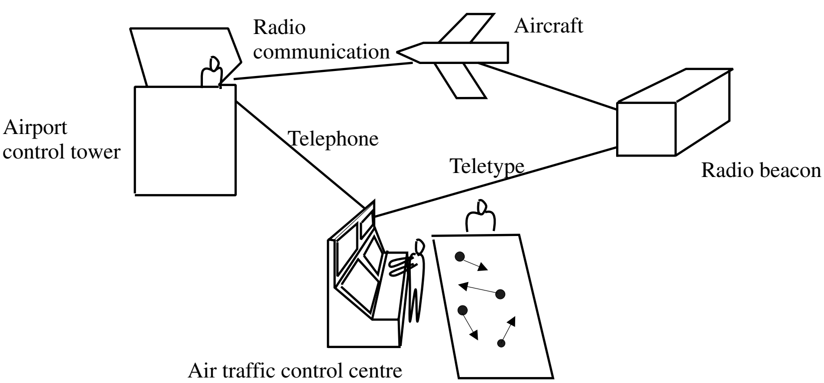
A barrier to greater market expansion of aviation was public perception. Although the wider public was enthusiastic about flying in the abstract, they had a fear of flying in practice, induced not only by safety issues, but also by the symbolic image of pilots as birdmen and daredevils, characterised by energy, courage and great physical dexterity. This symbolic image appealed to business and professional men, but not to the public at large. To deal with the public fear, airline companies set up publicity campaigns featuring well-known personalities travelling on aircraft (Bilstein, 1995: 95). Part of the campaign was to use female pilots. Nothing would convince the public of the safety of aviation as much as to see a woman flying an aeroplane (Corn, 1983: 75). Thus, by the late 1930s, flying had been purged of any association with danger. It was perceived as being less dangerous and people's fears were allayed. The new respectability was indicated by trip insurance costs. In 1940 the Equitable Life Assurance Society determined that air travel had reached a level of safety that justified a reduction in trip insurance from one dollar per unit of $5000 coverage to 25 cents, the same as for rail travel (Bilstein, 1995: 98). Once confidence in flying had been established, women were downgraded from pilot to stewardess.

The use of the DC-3 required changes in infrastructure, because its greater weight increased wheel pressure on the runway surface. In response, airports switched to concrete runways in the 1930s. Heavier aircraft also required longer runways for take-off and landing. The new runways triggered further innovations (Dierikx and Schot, 1998: 36). Because the landing of aircraft became more difficult on concrete surfaces, which had less resistance, landing flaps and special wheel brakes were developed to improve the braking capability of aircraft. The growth of passenger traffic also put pressure on existing terminal facilities. New airports required all operations to be scaled up to accommodate the new demands of commercial air transport (Douglas, 1995: 69). Despite their growing scale, airports remained insolvent operations. In order to deal with financial problems and stimulate airport expansion, revisions were made in federal policy. With the Civil Aeronautics Act of 1938, the federal government was officially allowed to build and provide airports (ibid.: 70).

Concrete runways increased the need for precision landing. Whereas previously the pilot could land on a grass field, he was now restricted to landing on the runway. To improve precision a new navigation system was developed, using radio technology. Approaching aircraft received radio signals from radio beacons around the airport, helping them to determine the right flight path: if the plane was on course, the navigator heard a constant tone in his headphones; if the plane deviated, the tone changed. In this early instrument landing system (ILS) the pilot still had considerable autonomy.

Another innovation was a response to the crowding of air space around airports, which posed threats to safety, and there were debates about stricter control of aircraft. Many pilots disliked the idea, because it reduced their autonomy. But a major crash in 1935 provided the incentive to create an air traffic control (ATC) system, which monitored airways and directed aircraft (La Porte, 1988: 217). The ATC system was operated manually. The air traffic controller had to keep track of flights within 50 miles of an airport, using a blackboard, a table map of local airways, a telephone and teletype (Figure 6.11). Aircraft emitted radio signals to radio beacons on the ground. With three stations it was possible to use time differences between the arrival of signals to calculate the position of the aircraft. On a map, flight paths of different aircraft were represented by small 'toys'. The ATC centre communicated flight paths to the control tower, which directed pilots via radio.

Aeronautical law had been a hotchpotch of publications and directives. The airline companies wanted clearer government regulation, because they were pushed and pulled in different directions. The late 1930s saw a general tightening-up of the rules. Pilots had to file flight plans and have them reviewed prior to take-off to prevent collision. Air traffic controllers were given more authority to give binding directions for landing routes. The concept of 'controlled airspace' was developed in 1936. The initial



Source: Gilbert (1973: 97); courtesy of Smithsonian Institution Press.

Figure 6.11 First-generation air traffic control system (mid-1930s) strangeness of this concept is shown by a comment in the *Washington Post* in October that year:

> 'The sky isn't the limit anymore. They've started limiting the sky. Not only limiting it but cutting it up into lanes and channels, with designated intersections and traffic rules and watchful policemen and fines for reckless drivers.' (cited in Heppenheimer, 1995: 122)

The Civil Aeronautics Act (1938) provided a codification of rules. It also created the Civil Aeronautics Authority, whose mission was to preserve order in the industry. It had the power to regulate airline tariffs, airmail rates, mergers and airline routes. It also created the Air Safety Board to regulate safety.

Dynamics in the Military Regime (1936–1940): War Preparations

In Germany, war preparations in aviation had a focus on tactical dive bombers (the Ju87 Stuka) and fighter aircraft (the Messerschmitt Bf 109). Within the German Air Ministry (RLM) there was an early understanding of the possibilities of new propulsion systems (Constant, 1980). In the second half of the 1930s, the RLM coordinated several innovative aero-engine projects (for example, turbojet, turboprop) and airframe developments (for example, the Me-262).

By late 1935 Britain had started a rearmament programme. The British strategy consisted of a mixed approach in which defensive fighters played a vital role next to bombers (Nahum, 1997: 312). R&D programmes were accelerated in three fields: radar, fighters and aero-engines. In the late 1930s, Britain developed radar technology (Eldridge, 2000: 5). By the outbreak of war a network of radar stations along the British coast was in place, acting as an early-warning system. Radar gave the British air forces a tactical advantage during the Battle of Britain, because it allowed the Royal Air Force to direct fighters to the required spot. Britain developed two new fighters: the Hurricane and the Spitfire (speed around 565 km/h), both fitted with Rolls-Royce Merlin piston engines. Prior to the use of radar, interception of bombers relied on 'standing patrols' of fighter aircraft, already flying in the air. These patrols were relieved periodically as their fuel became exhausted. In this context, low fuel consumption was a crucial performance dimension, guiding the development of aero-engines. Radar technology led to a change in thinking about air defence. Fighters could stay on the ground until enemy aircraft were spotted by radar. They then had to climb quickly to a great height from which to intercept approaching bombers. Hence the guiding principle shifted from fuel efficiency to engine power in order to give rapid climb rates. In the search for high-power engines three options competed (Nahum, 1997): (a) further improvement of piston engines, (b) development of two-stroke, fuel-efficient diesel engines and (c) development of jet engines. Most money eventually went to the first option, although the jet engine also received its share.

In America there was no sense of urgency to expand the air fleet. But new aeroplanes were on the drawing boards, in particular the B-17 Flying Fortress, a heavy four-engine bomber, with a cruising speed of 170 mph, capable of carrying 20,000 pounds of bombs over 2000 miles. When war broke out in Europe in 1939, America was inadequately equipped to face a world crisis. In August 1939, the American Air Corps had 800 planes, compared with 3750 for Germany and 1750 for Great Britain (Rae, 1968a: 108). The Navy added another 800 planes, but many of these were old biplanes. The United States had the best heavy bomber, the B-17, but only 23 of these were in service in September 1939. The worst situation was with fighters, which were inferior to foreign competitors.

Emergence of the Turbojet in the Interceptor Fighter Niche

Military planners and R&D managers in Britain and Germany became more interested in high-performance aero-engines, especially for interception fighters. The increased interest led to a support network for ideas by Whittle, von Ohain, Wagner and Muller.

In Britain, Whittle's company, Power Jets, tested a prototype jet engine in April 1937. It ran out of control, because designers did not yet know the characteristics of the complex engine. By late 1937 the ARC gave Whittle's company a small contract for further development and testing of his jet engine. In the 1938 tests, the Whittle engine reached 13,000 rpm, measuring a thrust of 480 lb (Gunston, 1997: 129). The test ended because the turbine nozzle fouled the rotor, and friction caused nine turbine blades to fail. But a new demonstration project in mid-1939 convinced the Air Ministry of the jet's potential (Constant, 1980). The status of the turbojet changed from long-term research to practical development. Increased government spending led to a proliferation of turbojet projects and British aero-engine companies jumped on the bandwagon (for example, Metropolitan-Vickers, De Havilland, Rolls-Royce). There was also renewed attention for the turboprop engine from the RAE in 1936. But the turboprop proved difficult to build and the RAE redirected resources to the turbojet.

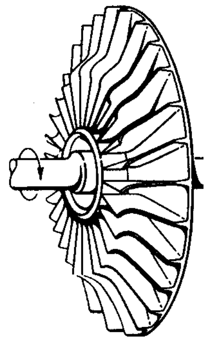
German turbojet pioneers linked up with existing aircraft firms. Wagner and Muller teamed up with Junkers aircraft and an engine was tested in 1938. Von Ohain linked up with Heinkel Aircraft in 1936, and a prototype first ran in 1937, giving a thrust of 250 kg (551 lb) (Gunston, 1997: 125). In 1939 an improved version produced a static thrust of 450 kg (992 lb), and Heinkel authorised construction of a jet aircraft. The world's first jet aircraft flew on 27 August 1939, reaching a speed of 450 km/h. Although the engine's combustion chamber overheated, the practicality of jet-powered flight was proven. The success triggered a proliferation of projects in Germany, coordinated by the RLM, involving all four major German aero-engine manufacturers: Daimler-Benz, Junkers Motors, B.M.W. and Bramo. The RLM also turned its attention to airframe development, approaching Messerschmitt in 1938 to make preliminary investigations into airframes for turbojet propulsion. The favoured design eventually became the Me-262.

The military niche provided the opportunity to investigate the three main components of the jet engine. Researchers and technicians were venturing into unknown areas of high temperatures, high airflow speeds, supersonic aerodynamics and high tensile material stress. There were very few technical models, guidelines or rules of thumb. Design was a 'dark art', in which designers used crude empirical formulae, slide-rules and tables. Engineers and technicians had to find their way through probing and learning.

Combustion was a problem, because a large amount of fuel had to be burned in a small chamber. To burn much fuel, combustion had to take place at high temperatures (about 500–600°C in early engines). This placed high demands on materials, which often became brittle. Material innovations were called for, leading to experiments with different metal alloys. Another problem was that the tremendous amount of energy proved difficult to control. Combustion often took place outside the combustion chamber, damaging the turbine blades. These problems posed a range of design questions, about the shape of the combustion chamber, fuel injection, mixing of air and fuel, and combustion control.

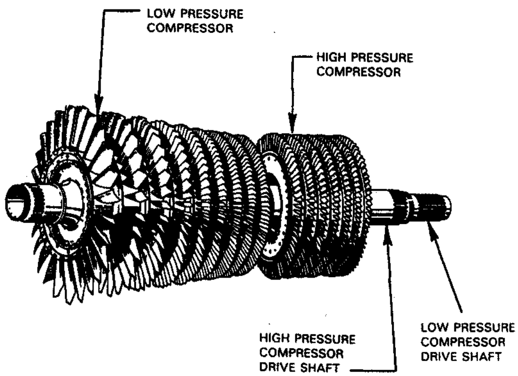
Another problem area was the turbine, which had the task of extracting energy from the hot gas leaving the combustion chamber, converting it into shaft power to drive the compressor. An important design issue was the shape of turbine blades, which should absorb energy from the gas, but not too much. A further consideration was that the gas should leave the turbine with the whirl removed, as any residual whirl would increase scrubbing on the jet pipe, losing energy and causing vibration. The blades should also be strong enough to withstand shockwaves when the gas flow reached the speed of sound (ibid.: 42). Blades should also be able to withstand severe centrifugal stress caused by the high-speed rotation, and high temperatures, which could make them brittle and cause them to break.

A third problem area was the compressor. The high combustion intensity required a strong airflow to bring in enough oxygen. The compressor sucked in air and raised it to a high pressure. There were two basic design options: centrifugal flow compressor and axial flow compressor. The centrifugal type



Source: Gunston (1997: 17, 18); courtesy of Patrick Stephens Limited.

Figure 6.12 Centrifugal compressor



Source: Gunston (1997: 17); courtesy of Patrick Stephens Limited.

Figure 6.13 Axial-flow compressor

comprises a rotating disc, carrying radial vanes. Air is drawn in near the centre and flung out between the vanes by centrifugal force, leaving it with very high tangential velocity. The air then passes through a surrounding diffuser, which converts the high speed into high pressure (see Figure 6.12).

The second option was the axial-flow compressor (so-called because the air flows in an axial direction parallel to the direction of rotation), consisting of a drum on which are mounted a large number of radial blades (see Figure 6.13). Each ring of blades is called a stage, and between the rotor stages are arranged fixed vanes, called stators, so that the air alternately passes between the rotor blades and the stator vanes. Each stage makes the air more compressed and hotter.

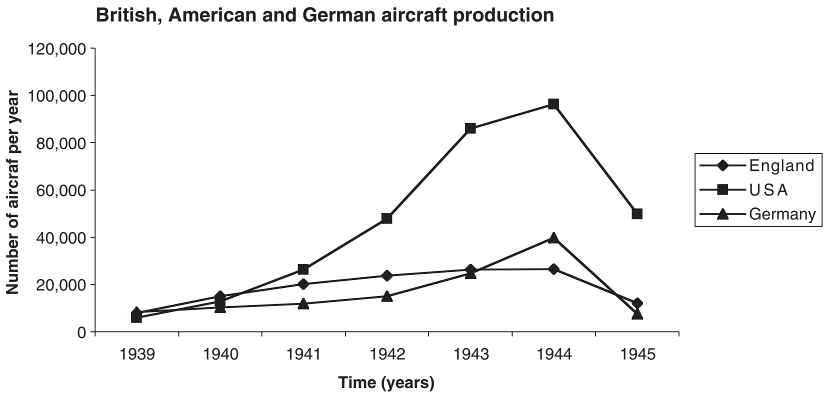
Both compressor types had problems. Although the basic design of axial-flow compressors was simpler, they were very difficult to construct with the materials and techniques of the late 1930s. Axial-flow compressors were prone to failure, because of vibration problems. Each blade is like a tuning fork, and any stalling of flow can cause vibration, severe enough to snap blades off at their roots. Turbojet pioneers such as Whittle and von Ohain chose the centrifugal type, which was cheaper, more robust and more predictable in behaviour (ibid.: 17). But in Germany there were also some projects that developed an axial-flow turbojet.

6.4 EXPANSION OF AVIATION AND THE BREAKTHROUGH OF TURBOJET AIRCRAFT (1940–1945)

Dynamics in the Military Regime During the Second World War

The Second World War was, of course, the overriding landscape development in this period. During the war, military aviation budgets were raised, creating more ‘space’ for technical innovation. While most money was spent on the improvement of existing technologies, turbojets also received more money. The development of new aircraft technologies was situated in the context of wider military strategies and their evolution over time. Early in the war, Britain was under attack from German bombers, and needed high-performance interceptor fighters for defence. Hence Britain invested in turbojet projects early in the war. Late in the war the tide had turned, and Britain needed bombers, while Germany needed interceptor fighters. Hence, Britain invested less in the turbojet by the end of the war and Germany more. Britain shared its jet engine knowledge with America in 1941. Although America did not produce jet fighters, they developed some of the most powerful jet engines by the end of the war. America paid less attention to radical innovations, and focused more on mass production of existing aircraft (Zeitlin, 1995), mainly transport aircraft and heavy bombers (B-17, B-24), which were eventually produced in greater numbers than in Britain and Germany (Figure 6.14).

Because of their emphasis on heavy bombers and transports, the American Army used the pre-war designs for four-engine aircraft and converted them to military use (for example, Douglas DC-4, DC-6 and the Lockheed Constellation). The use of these four-engine aircraft during the war demonstrated their reliability and performance.



Source: Data from Zeitlin (1995).

Figure 6.14 Aircraft production during the Second World War

The Development of Turbojet Aircraft in the Interceptor Fighter Niche

The early jet engines had high performance in thrust, but low performance in fuel consumption. Hence, the first application niche of the jet engine was the interceptor fighter. Although jet interceptor fighters could fly very fast and very high, they had only a limited range, and could not be used as escorts for bombers. Because speed was crucial for interceptor fighters, optimisation occurred on the dimension of thrust. The way to further increase thrust was to increase the combustion intensity, which required higher pressure and higher temperatures in the combustion chamber. To withstand higher temperatures, new heat-resistant materials were developed, for example, nickel-chromium alloys. Combustion remained difficult to control, and the fierce heat of the exhaust gases often damaged turbine blades. But nickel-chromium alloys helped to strengthen turbine blades. Engineers gradually began to understand the technical characteristics of the jet engine and the behaviour of the main components. This improved understanding led to major improvements in thrust performance. While the early Whittle and von Ohain engines produced 480 and 551 lb thrust, by the end of the war jet engines produced between 2000 and 4000 lb (Gunston, 1997).

Turbojet investments of different countries depended on general war strategies. Early in the war Britain invested substantially in jet engines, because they needed interceptor fighters as a defence against German bombers. Germany's military planners, on the other hand, made few investments in jet engines, because they needed tactical bombers to support their ground forces. By the end of the war, Britain was attacking Germany, and needed bombers and long-range escort aircraft. Investments in jet engines were decreased. Germany on the other hand, invested more in turbojets as defence against Allied bombers.

In Britain, the first test flight with a jet aircraft took place in May 1941, reaching a speed of 480 kmph. Although this speed was lower than the piston-engine Spitfire (565 kmph), it was promising for a test flight. After 1941 there was less pressing need for jet engines, but some development projects continued, all focused on centrifugal flow engines. The Gloster Meteor I went into service in May 1944. At the end of the war Britain had more powerful jet engines in development, for example, the Rolls-Royce Nene RB.41, which gave a 5000 lb thrust (Gunston, 1997: 136).

Germany had successful turbojet programmes before the war, resulting in two working engines, the Von Ohain (001) and the Wagner–Muller (006). But in 1940 German turbojet development was slowed down, because the RLM placed a prohibition on all research projects that could not be ready for production within a year, by which time the war should be won. For the next three years the development programme quietly progressed on a smaller scale at Heinkel, B.M.W./Bramo and Junkers Engine Division, but with little response from the chiefs of staff. German engineers came to focus more upon axial jet engines, because these had a smaller frontal area and could more easily be fitted under aircraft wings. BMW developed 003 axial-flow jet engines designed to power the Messerschmitt Me-262 aircraft design. This was a revolutionary design with swept-back wings, which made it possible for the aircraft to overcome shock waves at very high speeds. On a test flight in March 1942, both engines failed and the test was postponed. Hence, Messerschmitt turned to the Junkers Jumo-004 engines, larger and heavier than the 003, but more reliable. The 004-engine was made largely of steel, to reduce the dependence on strategic alloys. The sole source of chromium at the time was in British hands. But the use of steel reduced the engine's heat resistance, and was the reason why turbine blades were easily damaged. The steel jet engines had a short operational life. When they were later used in combat, they could only be used for 10–12 hours, after which period the engine had to be removed and completely rebuilt. With the 004 engines, the Me-262 made its first jet flight on July 1942. It entered pilot production in mid-1943 and was ordered into full-scale production near the end of that year. Although the 262 was built from poor-quality materials, it outperformed all other fighters of its time, with its high speed of nearly 525 mph (844 kmph). Nevertheless, it had two operational problems. Jet aircraft needed long runways for take-off, because jet engines are relatively inefficient at low speeds. Second, swept-back wings functioned well at high speeds, but created stability problems at low speeds. Only 1400 Me-262 aircraft were produced, but their performance impressed the aviation community and changed perceptions. By the end of the war, jet engines were seen as an important direction of future development.

America also developed turbojets, but did not apply them in fighter aircraft, mainly because their focus was on heavy bombers. American engineers were given the secret turbojet knowledge by Britain in 1941. Traditional engine manufacturers (Pratt & Whitney, Wright Aeronautical) were deliberately excluded from the development of the turbojet, allowing them to focus on piston engine production. Contracts were given to three outside steam-turbine firms (Westinghouse, Allis Chalmers and General Electric: GE) to develop both centrifugal- and axial-flow jet engines (ibid.: 143). GE in particular made good use of this opportunity and became one of the leading jet engine manufacturers by the end of the war. In 1944, GE put the centrifugal J-33 engine delivering a 4000 lb thrust, into production. GE also developed the axial J-35 engine, with similar thrust as the J-33, but smaller, although somewhat heavier and more expensive. The J-35 powered a range of American aircraft after the war. Of the established aero-engine manufacturers, Pratt & Whitney was able to make the jump to jet engines in the post-war years. But Wright Aeronautical stuck with piston engines and eventually went out of business.

Dynamics in the Civil Aviation Regime

While European commercial aviation declined markedly during the Second World War, American domestic passenger transportation continued to grow, providing a stimulus for airline companies (see Figure 6.10). Because American wartime production focused on heavy bombers and transport aircraft, American manufacturers would be well placed to produce large civilian aircraft in the post-war period and dominate the market.

Britain did not like this prospect, setting the stage for growing rivalry regarding the future of post-war commercial aviation (Benson, 2000). In 1942, the British government set up the Brabazon Committee to develop plans for the design of future civil aircraft, which could give Britain a chance to compete with American airlines. The committee recommended a leapfrog strategy, using Britain's advanced R&D capabilities to develop leading-edge technologies in a few large-scale civilian projects. One of these projects was to develop a civil jetliner, the Comet, exploiting Britain's advance in turbojet technology. Because the Brabazon Committee was packed with representatives from the British elite, its outlook was somewhat conservative and pre-war. Thus, the commercial focus was East, to the colonies, instead of West, to America. The design specifications were guided by this user perception. The first Comet was a fast, but small aircraft (carrying only 36 people), its limited range making it unable to cross the Atlantic Ocean.

As part of the rivalry in commercial aviation, Britain frustrated international expansion of American airline companies by denying them landing rights throughout the British Empire (ibid.). To overcome this barrier America organised an international conference in November 1944 to discuss the future of aviation, proposing an open skies plan. Britain opposed this plan, and called for the formation of an international authority with strong regulatory powers, restraining American airline operations. In 1945 the International Air Transport Association (IATA) was formed. The IATA limited competition by setting minimum rates and standards of service for international airlines. This gave some protection to British international airline companies.

6.5 EXPANSION OF CIVIL AVIATION AND THE INTRODUCTION OF THE TURBOJET (1945–1960)

Landscape Developments

After the Second World War, Allied powers and the Soviet Union occupied Germany and divided Europe in control spheres. The Soviet Union increased its political and military power in Eastern Europe and Winston Churchill warned about the descent of an 'iron curtain'. The communist putsch in Czechoslovakia in 1938 and the Soviet withdrawal from the shared administration of Germany heralded the beginning of the Cold War, an important landscape development, which influenced much of geopolitical and military thinking. The Berlin blockade (June 1948–May 1949) and the Korean War (1950–53) reinforced the animosity between America and the Soviet Union. Aircraft played an important role in the Cold War, especially as heavy strategic bombers guaranteed mutually assured destruction with nuclear weapons. This landscape development ensured continued military R&D spending in the aviation regime. This, in turn, allowed the jet engine to break out of the niche level, and acquire a prominent role in military innovation strategies. The jet engine was first developed and implemented in fighter aircraft, something which received a boost from the Korean War. As the turbojet was further developed, it was also used in bombers. After initial uncertainty the jet engine diffused widely in the military domain.

Another landscape development was the economic boom in post-war Europe and America. The growing economy stimulated the further rise of commercial aviation, and increasing trade contacts between Europe and America also boosted international aviation. Jet engines were initially perceived as unsuited for civil aviation, but strategic games between Britain and America eventually led to an exploration of market niches for jetliners in the growing international passenger market.

Ongoing Dynamics in the Civil Aviation Regime

The war had some stimulating effects on civil aviation in America. During the war American airline companies gained an advantage in operational experience over their European counterparts. Between 1939 and 1945 the British Overseas Airways Corporation (BOAC) flew only 65 million miles and carried 320,000 passengers (Benson, 2000: 30). American airline companies flew 3 billion passenger miles during the 1942–45 period. The Americans had more planes, more pilots and more experience. The war had also greatly stimulated American aircraft manufacturing, for example, the introduction of mass-production methods. American aircraft production was standardised and operated at low production cost. Furthermore, the building of heavy bombers and transport aircraft had created a knowledge base which manufacturers could use in the production of large civil airliners. Another effect was that the American government dumped many military aircraft on the commercial market, sometimes for as little as $90,000 (Heppenheimer, 1995: 114). This enabled the entry of small airline companies with cheap aircraft. Cheaply operating, non-scheduled airlines flew only when they had as many seats filled as possible. In response, major airline companies introduced cheaper flights in 1948. The competition led to a price war in 1949, resulting in ultra-low fares (ibid.: 127). This helped to make flying a means of mass transportation on domestic routes. Domestic air travel continued to grow rapidly and became a normal way of transportation (see Figure 6.2). In 1951, airline passengers exceeded train passengers for the first time. In 1956 more travellers between America and Europe went by aircraft than by steamship (Bilstein, 1995: 105).

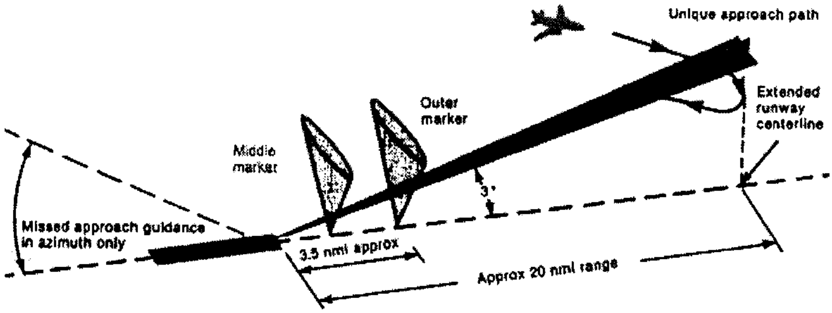
Having proved their performance during the war, four-engine aircraft entered commercial service (the DC-4, the Constellation and the DC-6). The aircraft were fitted with pressurised cabins, which allowed them to fly at greater altitudes. Because the planes were larger and flew higher, they were more comfortable. Scheduled non-stop coast-to-coast services started in March 1946. Because of the success in this market niche, aircraft designs were further stretched, leading to the Super Constellation (1950) and the DC-7 (1953). The DC-7 had enough range to offer a non-stop transatlantic service. Although the DC-7 could fly 110 passengers, its basic aircraft design was not very different from the DC-3. But even with these four-engine aircraft, flying at really high altitudes was a problem, even with supercharging. Thus there was a conflict between piston power, altitude and comfort. Jetliners linked up to this conflict in the 1950s, offering flights at very high altitudes without power loss.

International aviation also grew rapidly. With four-engine aircraft, airline companies began non-stop, coast-to-coast flights and launched new routes across the Atlantic and the Pacific. Global travel became a reality. Price wars on international routes were prevented by the 1945 cartel agreements that had been made in the IATA, keeping international fares artificially high. But in 1952, the IATA introduced tourist class tickets on transatlantic routes, which were approximately 32 per cent cheaper (Dierikx and Bouwens, 1997: 81). The term 'tourist class' indicates that airline companies were reorienting themselves towards a new consumer group: the leisure traveller.

In terms of symbolic perception, the experience of the Second World War created a new awareness of the ability of aviation to diminish barriers of time and distance. Air travel suddenly reached global maturity (Bilstein, 1995: 99). The expansion of flying in America also meant that more people began to see it as a 'normal' means of transport. Flying was no longer only for daredevils. There were also attempts to open up new markets and user groups, especially those with leisure and money for specialised vacations. The change in user groups was reflected in marketing, which came to emphasise exotic locations (Dierikx and Bouwens, 1997: 79).

As civil aviation grew, punctuality (flying on schedule) and safety became increasingly important guiding principles. Landing should be safe and also be possible in bad weather. Landing with low visibility required blind-landing systems of which two competed with each other. The first was the existing instrument landing system (ILS), using aircraft instruments and radio beams to create a three-dimensional glide path. A localiser transmitter provided the direction, and an inner and outer marker beacon gave precise indications of height. With specialised aircraft instruments, pilots could situate their aircraft with regard to an ideal approach path for landing (see Figure 6.15).

The second system, ground controlled approach (GCA), made use of radar technology developed in the military domain. Ground-based air traffic controllers would situate approaching aircraft in an ideal approach path with radar, giving pilots precise manoeuvring instructions via voice radio. The advantage of GCA was that it eliminated the need for special instruments inside the aircraft, and the need to train pilots to use them. On the other hand, GCA required a complex ground installation, and a large number of operators. The fundamental difference was who had the information necessary to ensure flight safety: pilots or controllers. GCA would mean a substantial decrease in pilot autonomy, which they opposed. The choice between both systems resulted in controversy: the



Source: OTA (1982: 94).

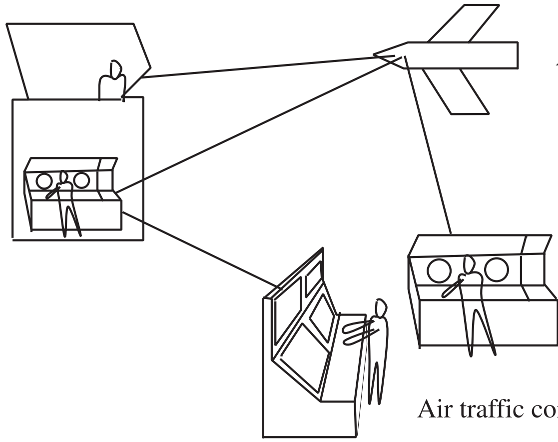
Figure 6.15 Constant-intensity glide path for ILS

Army, Navy and owners of private aircraft favoured GCA, while airline companies, commercial pilots and their representatives, the Civil Aeronautics Administration (CAA), favoured ILS. In the 1950s, ILS won the battle (see Conway, 2001 for a detailed analysis).

But this was not the end of radar. By the mid-1950s, pressure on airports was increasing as faster Super Constellations, DC-6 and DC-7 aircraft increased in numbers. In June 1956, two aircraft collided in mid-air, killing 128 people (Heppenheimer, 1995: 178). The collision pointed to the lack of long-range radar, which could track and 'space' aircraft as they flew through airways. In response, Congress passed the Federal Aviation Act in 1958. This created a new regulatory agency, the Federal Aviation Agency (FAA), which was charged with establishing an ATC system to maintain safe separation of all commercial aircraft through all phases of flight. In addition, the FAA assumed jurisdiction over all other aviation safety matters, such as the certification of aircraft designs and airline training and maintenance programmes. Radar was thus adopted by air traffic control for en-route control. It was also used for airport surveillance, to space aircraft together more closely during their approach to airports without compromising safety. The ATC approach system at airports thus developed into a hybrid combination of technologies. ILS was used for landing and GCA for en-route control and aircraft spacing. Both the control tower and the ATC centre had direct communication with the pilot (see Figure 6.16). Radar was used in air traffic control, and aircraft instruments for landing.

Wide Diffusion of the Turbojet in Military Aviation

After the war, governments cut back on military expenditures. But they maintained a moderate demand for aircraft, because the war demonstrated



Source: Gilbert (1973: 98); courtesy of Smithsonian Institution Press.

Figure 6.16 Second-generation air traffic control: the advent of radar

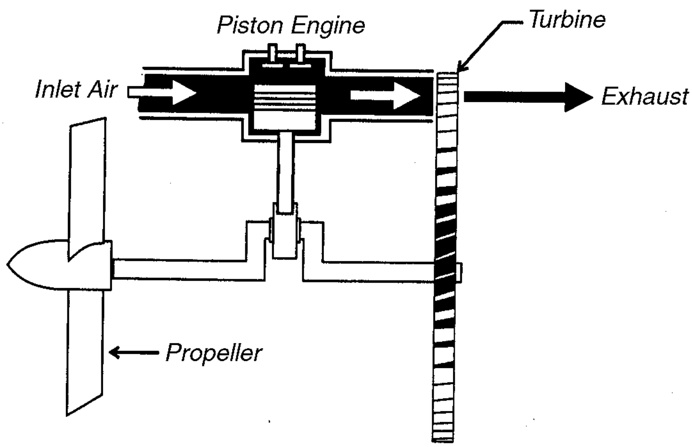
the military importance of air power. Governments wanted to keep their aviation industry alive, and push manufacturers to stay at the forefront of technical innovations. The military market increased during the Korean War and remained fairly high in the following decade. The Korean War was also a major boost for the production of jet fighters and bombers. The ensuing Cold War was a stimulus for the heavy long-range bomber, to retaliate in the event of a Soviet attack. In the mid-1950s strategic missiles emerged as a military option for strategic bombardment. Hence, the number of military bombers gradually decreased in the late 1950s.

The Second World War had changed perceptions in the aviation community about the jet engine. The shared perception was that jet engines had great potential for fighter aircraft, because of their high thrust. For this application, the engine's high fuel use was not a major problem. Not many American jet fighters were produced before 1950. But the Korean War marked a turning point. The confrontation with the Soviet MiG-15 fighter greatly surprised the Americans. With its powerful jet engines and swept-back wings, the aircraft performed at least as well as the American F-86 Sabre. Realising that the USSR also had jet fighter development programmes, America embarked upon a high-speed conversion programme to replace piston engines with turbojets. The conversion to jet engines required further changes in the aircraft, to overcome the problem of shock-waves at near-sonic speeds, which caused excessive drag. One innovation was the change in airframes: the adoption of swept-back wings. The other innovation was smooth skin airfoils, which reduced aerodynamic drag and the problem of kinetic heating (Miller and Sawers, 1968: 175).

The fighter niche provided resources for further development of the jet engine, resulting in substantial performance improvements. Different countries pursued different technical paths and innovation strategies. After the Second World War, turbojet technology was still in a state of flux. While Britain committed itself to the centrifugal jet, which was larger but more rugged, America tended towards the axial turbojet, which had a smaller frontal area, was more compact, but also more complicated (Dawson, 1995: 133). However, component innovations solved the technical difficulties of axial engines. As a result, axial jet engines emerged as the dominant turbojet design in the 1950s. During this process America overtook Britain as the leading country. Britain's post-war government regarded defence expenditure as anathema and showed no interest in the national lead in turbojet development (Gunston, 1997: 151). Despite the government's disinterest, Rolls-Royce was one of the best manufacturers of jet engines. Its centrifugal Nene engine was very well engineered and powerful, and delivered a 2270 kg thrust (5000 lb). Its sale to the Soviet Union in 1946, caused an international diplomatic conflict between Britain and America (Engel, 2000). America wanted to have the best jet engines in the world, because air power had become the central guiding principle in American military thinking. American engineers acquired and improved turbojet knowledge through reverse engineering of British engines, industrial espionage in British institutes and through fundamental research at the National Advisory Committee for Aeronautics (NACA) laboratory (Dawson, 1995: 135). American engineers gradually took the lead in turbojet development. By mid-1948, Pratt & Whitney produced a redesigned Nene engine, the J-42, delivering up to 8750 lb thrust with afterburner. Boeing, which was exploring the option of jet-powered bombers, asked Pratt & Whitney to design an engine that would deliver a 10,000 lb thrust. This became the J-57, a twin-spool and dual-rotor axial-flow compressor, which allowed low fuel consumption. Production of the J-57 began in 1953. The engine was widely used to power fighters, bombers, and even the early Boeing 707. In 1952, GE began a design for a new turbojet, offering enough thrust for Mach 2 speeds. This resulted in the J-79, which used variable stators and was first run in 1954. The J-79 matched the J-57 in thrust, but was lighter, smaller and less bulky, offering reduced drag and higher speed (Heppenheimer, 1995: 200).

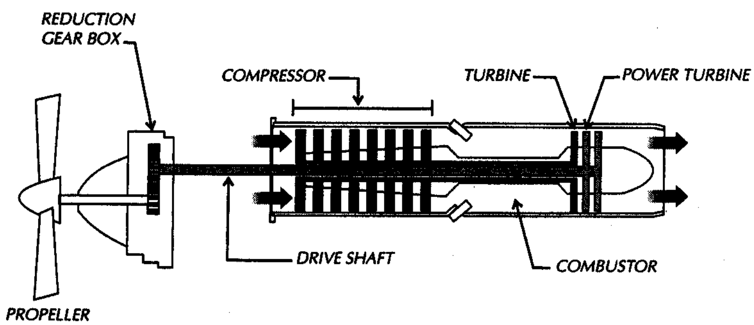
The tremendous performance improvements were achieved by innovations in different components of the turbojet. These innovations build on advancements in aerodynamics and material science. Material scientists mixed hundreds of metals to find alloys, which could withstand increasingly high temperatures (ranging up to 1100°C in the 1960s). One improved component was compressors, which formerly suffered from stall, that is, early stages of the compressor suck in more air than later, high-pressure stages can swallow; when a compressor stalls, the airflow pushes forward violently, sometimes causing blades to break off. One solution for the stall problem was Pratt & Whitney's twin-spool design, applied in the J-57 engine, where two compressors were mounted in tandem, each driven by its own shaft and turbine. The low-pressure compressor spool could be made to spin at the proper rate to avoid choking the high-pressure spool. Another solution was worked out by GE for the J-79 engine: the variable stator (Gunston, 1997). Varying the angle settings of stator vanes, stage by stage, could adjust the volume of airflow and prevent too much air from coming in. Another innovation was to add a fan in front of the compressor, and have it powered by the turbine. This fan, an array of rotating blades, would blow air past the rest of the engine, adding it again at the back along with the hot jet. This airflow is called the 'fan flow', or bypass flow, and the ratio of the fan flow to the core airflow is the bypass ratio (BPR). It was found that the so-called 'fanjet' or 'turbofan' offered better fuel economy, delivered more thrust and produced less noise. The engine was called a ducted fan or turbofan. Early fanjets had small fan blades. The Rolls-Royce Conway engine (1952) had a BPR of 0.3 (ibid.: 27). But the span of fan blades increased over time, and in the 1970s fanjets with a BPR of 5–7 were produced. A second component, which was improved, was the turbine. Better blade designs and new heat-resistant alloys improved its performance. A third innovation was after-burning. To use leftover oxygen in the combustion chamber, this device injected additional fuel in the jet pipe behind the turbine. The after-burner could increase thrust by 50–90 per cent, but at the expense of more fuel (ibid.: 53). Hence it was used mainly for short periods of time, for example, during take-off or in combat.

For bombers, where range and fuel efficiency were important performance criteria, there was more uncertainty about the engine of the future in the late 1940s and early 1950s. It was not known whether future bombers would be propelled with improved piston engines, turboprop engines or jet engines. The compound engine was an improved piston engine, developed in the late 1940s by the Navy to deliver up to 20 per cent more power (Heppenheimer, 1995: 128). It was a hybrid form of piston engine and gas turbine, which promised to combine increased power and better fuel economy. It went one step further than the turbocharger by having the gas turbine actually deliver power to the propeller shaft (see Figure 6.17). Introduced in 1950, the compound engines came to be used in bombers and commercial planes. The compound-engine bombers redefined what it meant to be a first-line bomber, with performance requirement raised to 10,000 pounds of bomb load for 10,000 miles (ibid.: 145).



Source: Heppenheimer (1995: 129); courtesy of John Wiley & Sons.

Figure 6.17 Compound engine



Source: Heppenheimer (1995: 146); courtesy of John Wiley & Sons.

Figure 6.18 Turboprop engine

The second engine was the turboprop, where a gas turbine drove the propeller (see Figure 6.18). The technology was very similar to the jet engine, but the focus was on the rotational power of the drive shaft rather than on the thrust of exhaust gases. The turboprop had been investigated in the 1930s in Britain, but its development was shelved in favour of the jet engine. But in 1948 the turboprop engine was introduced in the British Vickers Viscount, a passenger aircraft. The turboprop seemed to offer the best of two worlds: higher speed than piston engine aircraft and better fuel economy than jet engines (Gardiner, 1984: 130).

With longer range specifications for bombers, more-fuel-efficient engines were needed. Hence, ideas for future bombers concentrated on turboprops. In 1946 Boeing carried out research for the B-52 bomber, and seriously considered the use of six turboprop engines. But the idea was abandoned because of concerns about high stress in the connection between the engine and the propellers (Heppenheimer, 1995). Hence, Boeing engineers opted for jet engines in 1948.

Turbojets were the third kind of engine. Their great thrust offered high speed and could power large aircraft. But their high fuel use would limit the range. The variety in engines caused uncertainty in the bomber market niche. The potential of jet engines was eventually strengthened because of a complementary innovation: refuelling in the air. If jet bombers could be refuelled in the air, their high fuel consumption would be less of a problem and their range could be extended. Tests showed that jet engines could power very large bombers. Because the range problem was alleviated, Boeing engineers abandoned the turboprop for the B-52, the new long-range, heavy bomber, and opted for jet propulsion in 1948 (Rae, 1968a: 176). The B-52 first flew in 1954, and entered into service in 1955. The jet-powered B-52 was important on a symbolic level, because it showed that jet engines could power very large aircraft. In the aviation community this led to a change in perception, and jet engines came to be seen as powerful all-purpose propulsion engines.

Jet bombers emerged in tandem with tanker aircraft. Because piston engine tanker aircraft were too slow, jet-powered tankers were needed. Boeing built such a tanker aircraft in 1952, bearing in mind a dual-use strategy. Boeing would also try to sell the design as a jetliner to the airline companies. In America, heavy bombers and tanker aircraft thus formed stepping-stones towards a civilian jetliner. Britain followed another route as we shall see in the next subsection.

Exploring Market Niches for Jetliners in Civil Aviation

Between 1948 and 1955, the variety in engines caused much uncertainty in civil aviation, where cost was the most important selection criterion. This was especially the case in America, where private airline companies were involved in strong competition. In Britain, cost was a somewhat less important criterion, because BOAC was a state-sponsored airline company. Because of its high fuel consumption, the shared perception in America was that the jet engine would not be cheaper than piston engines. The turbojet engine was slow to enter American civilian aviation. The common view in the US airline industry was nicely expressed by Ralph Damon, president of Trans-World Airlines (TWA) in 1950: '"The only thing that is wrong with jet planes of today is that they won't make money"' (cited in Heppenheimer, 1995: 152). The American aviation industry chose a wait-and-see attitude with regard to the new jet and turboprop engines. The American industry could afford to adopt this strategy, because it dominated the aviation market, with its lead in airframe development and the new four-engine aircraft. Britain, on the other hand, was far behind in postwar civil aviation. Therefore Britain adopted a leapfrog strategy, as suggested by the Brabazon Committee during the war. As part of this strategy, British aircraft manufacturers, sponsored by the government, aimed to implement the jet and turboprop engines in new kinds of aircraft. The turboprop engine seemed to offer higher speeds than piston engines and better fuel economy than jets. British manufacturers introduced the turboprop in the Vickers Viscount (1948). This passenger aircraft was fairly popular in Europe, where aircraft flew shorter routes than in America. In 1950, 20 turboprop aircraft were ordered (Miller and Sawers, 1968: 183). The turboprop conquered market niches in the short- and medium-range markets, where it competed with piston engines.

Britain also took the lead in turbojet development, hoping to break America's grip on commercial aviation by using its technical advantage in gas turbines. Britain also paid some attention to the development of jet bombers, but far less than in America. Although the British Canberra bomber enjoyed some popularity, and was even exported to America, the British government decided to focus on commercial aviation. Following the Brabazon plan, the British government sponsored the development of a civil jetliner, the Comet. It did not matter whether the Comet would be expensive in operation, because the British state sponsored BOAC. The government also provided the manufacturer, De Havilland, with advantageous lending schemes to develop the plane, which began test flights in 1949. The early Comet was a small and fast aircraft, designed for the colonial market, that is, for senior civil servants and colonial businessmen. Its engines made it inadequate for long-range routes, for example, across the Atlantic. When the Comet entered commercial service in 1952 it amazed the world with its performance and was popular with customers. It had a revolutionary new design, with no obvious means of propulsion, as its four jet engines were built inside the wings, which were moderately swept back. It provided passengers with a quiet, smooth ride, high above the weather, reducing turbulence and increasing flight comfort. The Comet was also faster, raising flying speeds to 470–500 mph (750–800 kmph). Instead of a 40-hour flight to Johannesburg, the Comet could fly from London to Johannesburg, via Rome, Beirut and Khartoum, in 23 hours. The Comet's operating costs were nearly three times as much as the DC-6, but it flew with nearly every seat filled, and BOAC found that it was actually making money (Heppenheimer, 1995: 156). Because of its success, the Comet was redesigned to enter other markets. The Comet-2 was larger and used more powerful jet engines, the Rolls-Royce Avon-2, capable of flying an aircraft non-stop across the Atlantic. By May 1953, De Havilland had orders from around the world for 50 Comets and was negotiating for 100 more. It seemed that Britain had a good chance of becoming a major player in global aircraft manufacturing. But in 1954 the Comet dream was torn apart by a string of fatal accidents, caused by metal fatigue, an unfamiliar phenomenon at the time. The Comet had square windows instead of round ones, and metal fatigue occurred around the edges, leading to cracks in the airframe. It would take years to redesign and re-certify a new Comet. By then Britain would be out of the game.

The Comet was a wake-up call to American aircraft manufacturers. Both Boeing and Lockheed had been studying jet concepts since 1947 (Miller and Sawers, 1968: 183). Boeing had developed jet bombers and jet tankers for the military, which allowed it to build up valuable design experience. Douglas Aircraft made some preliminary designs for jetliners, but it adopted a wait-and-see attitude, letting Boeing carry the first-mover risks. Douglas had a dominant position in the civil aircraft market and thought it could afford this strategy. Boeing was strong in the military market, but a weakling in the commercial industry. In the early 1950s Boeing realised that the market for military jet bombers had its limits, because cruise missiles were expected to replace strategic bombers in the future. With the prospect of shrinking military markets, Boeing aimed to diversify to civilian markets. Therefore Boeing had exploratory talks with American airline companies about new jetliner designs, but they showed little interest. American airline companies were hesitant about operational costs, because jet engines were perceived as expensive gas-guzzlers with a limited range. Refuelling in the air might be possible for strategic bombers, but it was a not an option for civilian jetliners. Experience with the Comet had shown that fuel costs were more than twice as high as for piston-engine airliners (ibid.: 177). There was also doubt about the durability of jet engines, which was still a novelty. The high temperatures at which it ran, suggested that it could not be as reliable as piston engines. Experience with military jet aircraft did not allay the doubts, because civilian flying put more stress on the engine. On the other hand, the high speed and modern image of jetliners had a strong appeal to passengers. The jet's popularity might increase the plane's load factor. The Comet flew with an almost full complement of passengers, and jetliners offered more comfort, because they could fly high above weather, thus reducing turbulence and airsickness. Also, the higher speed meant that more flights per year could be made, increasing its yearly productivity. In sum, there was uncertainty and hesitation. Airline companies held each other in a 'cartel of fear'. No one wanted to take the first step, because of the risks involved. So long as no company introduced jets in the American market, no company needed to buy them.

It took a pioneer to break this inertia, and Pan American did just that. At the height of the Comet's success, Pan Am ordered three of them in October 1952, wanting to use them in the market niche of long-distance intercontinental routes, flying non-stop across the Atlantic. After the Comet accidents in 1954, Pan Am negotiated with Boeing and Douglas about possible jetliners. While Douglas Aircraft had adopted a wait-and-see attitude, Boeing had followed a dual-use strategy. For the market niche of jet tankers, Boeing had developed a new design, the Dash-80, which it could easily convert to a civilian jetliner. In May 1954 Boeing presented its prototype Dash-80, which convinced the Air Force to order 29 jet tankers. The Air Force thus paid for the design and also for the tooling and production equipment, much of which would carry over to the civilian jetliner. The Air Force order gave Boeing enough financial space to convert the design to the civilian Boeing 707, which it could offer in response to Pan American's demands. Stimulated by Pan Am's interest in jetliners, Douglas Aircraft also speeded up the development of their airliner, the DC-8. In October 1955, Pan Am ordered 20 Boeing 707s and 25 DC-8s. Once Pan Am had crossed the threshold, other airline companies followed quickly, for fear of being left behind, thus creating a bandwagon effect. This made the president of Delta Airlines utter a cry of despair in 1956:

> 'We are buying airplanes that haven't been fully designed, with millions of dollars we don't have. We are going to operate them off airports that are too small, in an air traffic control system that is too slow, and we must fill them with more passengers than we have ever carried before.' (cited in Heppenheimer, 1995: 170).

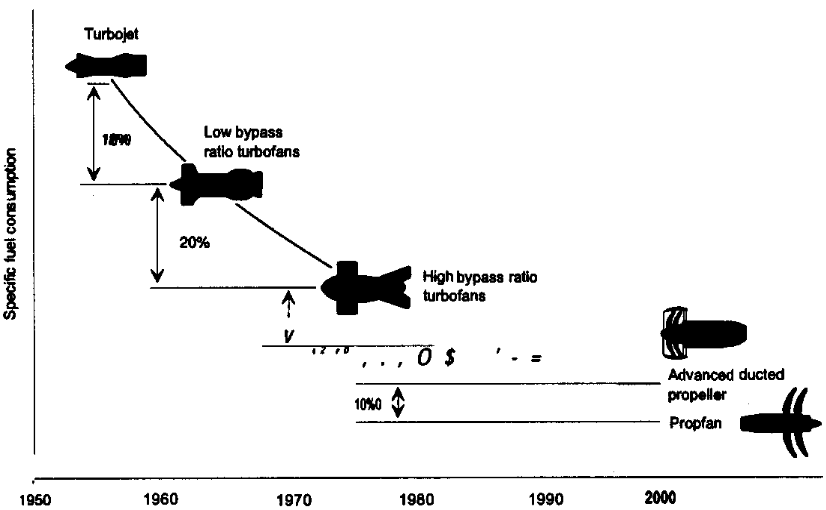
Boeing was able to acquire more customers than Douglas, and by the end of 1956 the 707 outsold the DC-8 by three to one. The first Boeing 707 was taken into service in 1958. It raised cruising speeds to 550–600 mph (885–965 kmph), accommodating up to 181 passengers. With its range of about 3000 miles, it became the workhorse of the world's airlines. The Boeing 707 embodied a major change in civilian airframes, namely swept-back wings. While this worked well at high speeds, it created instability at low speeds during landing and take-off. Therefore, additional innovations were made in wing flaps, tail and rudder configurations (Rae, 1968a: 210).

6.6 REPLACEMENT OF PISTON ENGINES AND WIDER IMPACTS (1960–1975)

American jetliners were first introduced in the long-distance market niche, where its advantage of high speed made the greatest difference. The further diffusion of jet aircraft in civil aviation was gradual. Civilian aviation was a heterogeneous domain, consisting of many market niches, for different routes and passenger loads.

In civil aviation, fuel efficiency became an important guiding principle for jet engine development, because it influenced both the range and the operating costs. The biggest contribution to fuel savings came from the adoption of fan-type engines. A new light and strong material, titanium, allowed the construction of ever-larger fans, which could stand the severe strain from centrifugal forces. The larger fans with higher BPRs enabled a further increase in thrust. Pratt & Whitney's JT-3D fanjet (1959) increased take-off thrust from 13,500 to 17,000 pounds. Fuel economy also improved markedly and noise was reduced by 10 decibels (Heppenheimer, 1995: 190). In subsequent years fans were further enlarged, resulting in turbofans with high BPRs (5–7). The introduction of higher BPR fanjets resulted in marked increases in fuel efficiency. Figure 6.19 gives an impression of relative improvements.

Jetliners were popular with passengers, because they were modern, fast, lacked the vibrations from piston engines, and could fly comfortably above



Source: OTA (1994).

Figure 6.19 Fuel-efficiency trends in aircraft engines the weather. Jetliners thus solved a long-standing problem in civil aviation: turbulence and airsickness. The adoption of jetliners was initially driven by the enthusiasm of the general public and by bandwagon effects among airline companies (fear of falling behind). Early jetliners did not enjoy high profits, and it took airline companies several years to reap the benefits of jet aircraft. Indeed, in the first years of the jet age, profits from airline companies actually decreased and were sometimes negative (see Table 6.1).

One reason for the lower profits were the high purchase costs: the purchase price of jets was almost three times higher, with a jetliner costing about $4 million per plane, while the DC-7 was $1.5 million (Heppenheimer, 1995: 163). Another reason was that the load factor decreased in the early years. Because jet engines were more powerful, aircraft could be enlarged, leading to a doubling of the average number of seats per plane. While average seats per mile doubled from 43.2 in 1955 to 86.5 in 1965, the load factor declined from 62.2 to 56.1 per cent (Tillinghast, 1966).

But after some years of experience, airline companies found that jetliners saved costs on several dimensions. Many of the cost savings were not foreseen in the early days of jetliners. Although jet engines used a lot of fuel at speeds below 400 mph, their efficiency increased at higher speeds. Jetliners also flew at higher altitudes, with less drag. Jets could also use cheaper fuel, which helped to offset the effects of higher fuel consumption (Miller and Sawers, 1968: 187). Furthermore, continued technical improvements (especially fanjets) increased the jet's performance, increasing fuel efficiency and reducing fuel costs. A major drop in operating costs came from economies of scale. The same number of crew could serve more passengers, and the higher speed meant that fewer crew were needed on long journeys. The higher speed also meant that more flights per year could be

Table 6.1 Operating profits from carriers of the International Civil Aviation Organization

	1955	1956	1957	1958	1959	1960	1961	1962	1963	1964	1965
Operating revenues (billion $)	3.0	3.5	4.0	4.1	4.8	5.4	5.8	6.6	7.2	8.2	9.3
Operating profits (million $)	78.0	84.0	–41.0	15.0	105.0	70.0	–118.0	97.0	326.0	601.0	890.0
%	2.6	2.4	–1.0	0.3	2.1	1.3	–2.0	1.5	4.5	7.3	9.6

Source: Tillinghast (1966: 3).

made, increasing its yearly productivity. Another large source of savings was maintenance (ibid.: 209). Because jet engines had fewer moving parts than piston engines, they had fewer breakdowns, resulting in savings on maintenance. It was only after some years of experience that airlines extended the time between one overhaul and the next. While the time interval for overhaul was 2000–2500 hours of service for piston engines, this was extended to as long as 8000 hours for jet engines (Rosenberg, 1986: 25). As a result of these changes, by 1963 a period of profitable operations began in the jet age.

As a reaction to lower load factors, airline companies tried to attract new user groups. They increasingly targeted the leisure market, citing tourism and exotic trips in their advertising and marketing. They also lowered airfares. To further stimulate the international long-distance market, IATA airline companies introduced economy class in 1958, as a new cheap ticket 20 per cent below tourist class (Dierikx and Bouwens, 1997: 83). This stimulated a shift in international air transport from elite customers to mass transport. The year 1956 was the first in which more people crossed the Atlantic by air than by ship (ibid.: 83).

In the market niche of long-distance transport, jetliners replaced the four-engine piston aircraft relatively quickly. For short- and medium-range routes, however, jet-powered aircraft initially had high costs and encountered strong competition from turboprop aircraft. Hence, the diffusion of jetliners and the replacement of piston engine aircraft followed a complex dynamic. This was also due to the existence of many different market niches, with different range and payload. Jets, piston engines and turboprops had a different appeal for the different market niches, often divided in terms of seat capacity. Competition between the three engine types varied for different market niches (Bonaccorsi and Giuri, 2000). The long-distance market niche with 120–180 passengers had been conquered rapidly by early jetliners. The 90–120-seat market niche was briefly entered by turboprops, in particular the Vickers Vanguard in 1960. But new jetliners were developed, with more fuel-efficient engines. Boeing scaled down its initial jetliner, resulting in the Boeing 727, introduced on medium-range routes in 1964. Because the 727 had low-speed landing and take-off performance it could be operated on smaller airports than the 707. The 727 became one of the best-selling aircraft in the world, enabling jets to be used on medium-range routes. In the 51–90-seat market niche, turboprop engines were used earlier than jet engines. The first turboprop airliner was the 60-seat Vickers Viscount, which first flew in 1948. In 1955 Fokker launched the F-27 turboprop, and in 1957 Lockheed introduced the Electra. Jet aircraft entered the medium-range segment in 1959 in the form of the purpose-built French Caravelle, which became fairly popular in Europe. The success of the

Caravelle encouraged the development of small jets in the 51–90-seat, market, including the Fokker-29 and the DC-9. The turboprop and jet engine continued to compete in this niche market during the 1960s and 1970s, with the jet gradually winning ground. In the markets with smaller seat capacity, the turboprop was more popular than the jet engine. In these markets, piston engines were not replaced by jets, but by turboprops. The only exception in this market was the German four-seat jet-powered VFW-614 which was launched in 1972. Between 1972 and 1978, only 19 aircraft were sold (ibid.: 855). In the markets with less than 30 seats, piston engines were gradually replaced by turboprops, but held out for a long time.

In sum, the diffusion of turbojets in civilian aviation was not only gradual, but also incomplete. The transition was not just from piston engines to jets: turboprops formed a third type of engine, which complicated the transition process. Thus, the transition was characterised by increasing variety in terms of markets, performance requirements and technologies (see also Frenken et al. 1999).

At the higher end of the market a new aircraft design was pioneered. The growth in long-distance, high payload routes opened up opportunities for new and larger aircraft. Expanding markets and technology co-evolved. The Boeing 707 was not only scaled down, but also scaled up, leading to a second generation of jetliners, the jumbo jet. In 1969 the Boeing 747 was introduced, representing a jump in passenger numbers (450–500), range (about 6000 miles), speed (640 mph, 1030 kmph) and altitude ceiling (45,000 feet). To power the 747, designers relied on a new jet engine, the high by-pass turbofan, with fan blades nearly 2.5 metres wide. This engine was quieter, cleaner, more fuel efficient and more powerful. The engine delivered more than 15,650 kilos of thrust, producing core temperatures above 1100°C.

The diffusion of jetliners was accompanied by changes on many dimensions of the socio-technical regime. The 747 was not just a technical breakthrough, but also an economic one. Fuel-efficient high BPR turbofans and scale economies allowed for a 30 per cent reduction in operating cost per seat mile. The combination of lower tariffs and rising incomes made flying truly a mass phenomenon in the 1970s.

Jetliners and jumbo jets boosted international tourism, which grew by 10 per cent per year between 1950 and 1971. In this period receipts from international tourism went up from $2 billion to $20 billion (Lonati, 1972: 12). Other developments also contributed, for example, rising purchasing power, labour laws providing more leisure for workers, expansion of tourist facilities, marketing campaigns and so on.

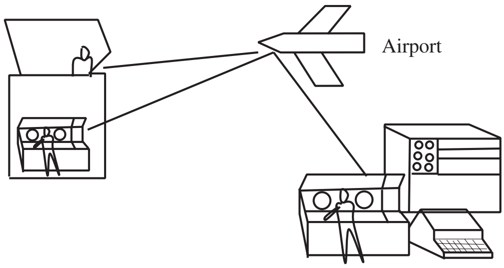
The new jetliners needed longer runways for take-off, because jet engines were relatively inefficient at low speeds. Jetliners were also heavier than piston engine aircraft. The original 747 was more than twice as heavy as the 707. Hence, runways had to be strengthened. Furthermore, the surface of the runways had to be adapted in order to resist the high temperatures of exhaust gases (1000–1100°C).

Jetliners also required that pilots, accustomed to piston engine planes, had to be retrained. The piston engine planes, with their straight wings, normally sat on their landing gear at a slightly positive angle of attack. When they reached take-off speed, they more or less flew themselves off the ground. The jets, with their swept-back wings, not only had much higher take-off speeds, but also had to be powerfully 'rotated' (pulled up with their elevators) at take-off speed. Pilots had to learn to fly 'by the numbers'.

The blast from early jet engines caused a lot of noise around airports, and the use of jetliners in cities such as Paris, London and New York placed noise on the political agenda. In 1962, the American Supreme Court decided that airport operators could be held liable for damage resulting from aircraft noise. Therefore, some airports decided to set noise limits for aircraft and installed noise monitoring systems (Smith, 1989: 20–21). But noise problems continued, and local residents came to see aviation as 'technology out of control'. The noise problem triggered further technical changes in jet engines. Noise became an important design criterion for aero-engines in the 1960s (van de Poel, 1998: 242). As a happy coincidence the fan-type engine was not only more fuel efficient, but also quieter. Although the turbofan reduced noise substantially, it did not entirely solve the noise problem, which continued to plague civil aviation.

The growth of aviation created new problems, as skies became more crowded. One response was to define segments of airspace, called sectors. With regard to air routes, a new national air control system was introduced in America in 1961, providing routing direction and radar advisories along three tiers of airways: (i) lower level from 1200 to 14,500 feet, (ii) intermediate airways from 14,500 to 24,000 feet and (iii) high altitude jet ways above 24,000 feet. By 1964, the three-tiered airways had given way to the present two (La Porte, 1988: 234).

The growth of aviation also increased the stress on ATC systems. The main stress on these systems arose from the increasing difference in aircraft speeds. Around 1950, cruising speeds of aircraft were rather close together, at about 370 kmph. By 1959, however, speeds ranged from 400 to 950 kmph. This complicated the task of air traffic controllers, who were still calculating the position of aircraft by hand. The control of the airspace around airports had to be faster, more precise and cover larger distances. With the growth of air traffic capacity, these requirements led to an increasing volume of bookkeeping functions and data-processing requirements (Gilbert, 1973: 98–99). The 'traditional' methods increasingly ran into problems. Computers were



Source: Gilbert (1973: 100); courtesy of Smithsonian Institution Press.

Figure 6.20 Third-generation air traffic control: automation, computers

introduced to deal with these requirements, signalling the beginning of the third-generation ATC-systems (see Figure 6.20).

6.7 CONCLUSIONS

In this concluding section I shall follow the case-study protocol, described in Chapter 3 (Section 3.5). I shall test first the multi-level perspective and then the different patterns.

Testing the Multi-level Perspective

In Chapter 3 (Section 3.5) two questions were formulated as proxies to test the usefulness of the multi-level perspective. How did niches emerge in the context of the existing regime? To which ongoing processes at the regime and landscape levels did the niche link up? Both questions will be addressed below.

An important claim of the multi-level perspective is that new radical technologies emerge in specific niches in the context of the existing regime and landscape. This claim was substantiated in the case study. The gas turbine was first developed in the regime of electricity generation (early twentieth century), where the focus in technological development was on the rotational energy of the drive shaft. This guiding principle remained strong when aviation engineers first thought about using gas turbines in aircraft in the 1920s. They proposed to replace the piston engine and use the rotational power of gas turbines to drive propellers (turboprop engine). The idea was examined in the 1920s, but not tried out in practice. The first practical niche for the gas turbine emerged within the piston engine propeller regime, as a response to the engine-altitude problem (less oxygen at high altitudes hindered combustion). The gas turbine was used in existing piston engine aircraft as an auxiliary device in the form of turbo-superchargers. Superchargers and turbo-superchargers were developed in the 1920s to increase the supply of oxygen. In turbo-superchargers, the focus was still on the rotational power of gas turbines. It was not until the early 1930s that some academic pioneers began to focus on the *thrust* of exhaust gases of the gas turbine, which led to ideas about the jet engine. These pioneers had great difficulty attracting the interest of actors in the established aeronautic community. Because the promise of the jet engine was not immediately recognised in the wider aviation community, it remained a 'hidden novelty' in the early 1930s. In the context of the Second World War, the jet engine was gradually developed and used in interceptor fighters.

The other claim of the multi-level perspective was that the regime should not just be seen as a barrier for radical novelties. Established regimes may be stable, but they are not inert. In socio-technical regimes and landscapes there are ongoing processes which may provide windows of opportunity to which niche innovations can link up. The case study did indeed show that the aviation regime was not static in the 1930s and 1940s. The gas turbine and jet engine linked up with several ongoing processes. This happened in two ways.

First, there were some developments, which preceded the jet engine and created a platform for its emergence. The airframe revolution of the early 1930s, for instance, resulted in stronger and smoother aircraft. The airframe revolution created the conditions for the turbojet. Without aircraft being stronger, the jet engine could not have been used because it would have torn them apart. The airframe revolution also helped to create the basis for the 'presumptive anomaly' which some academic pioneers identified (Constant, 1980). Assumptions about higher aircraft speeds in the future led some pioneers to conclude that the propeller formed the main obstacle. This presumptive anomaly triggered the search for new propulsion options, which resulted in the jet engine. Another example of building upon previous developments was the use of gas turbines for turbo-supercharging in existing piston engine aircraft. This niche generated a great deal of technological knowledge and experience useful to the later turbojet.

The second way of linking up was mutual reinforcement between the jet engine and ongoing trends and developments. One trend was the increasing speed of aircraft and the high-speed problem. Since the 1930s, aircraft have flown ever faster in order to offer a better service than their competitor: the train. Especially on long distances, the time difference between plane and train could be substantial. Rich passengers (businessmen, high-ranking officials) were willing to pay extra for high-speed travel. But propeller-powered aircraft suffered from the high-speed problem. As propeller tips approached the speed of sound, a 'compressibility burble' developed, which increased drag. This problem was overcome with the jet engine, thus allowing the continuation of the trend towards higher speed. Another trend was flying at greater altitude, preferably above the weather because this reduced turbulence and airsickness. In the late 1930s, four-engine aircraft with pressurised cabins were designed which could fly higher. But at higher altitudes, the power of piston engines and the grip of propellers were weakened (the high-altitude problem). Supercharging somewhat increased the altitude level, but it also had its limits. Thus, there was a conflict between motor power and more comfort, and it was this conflict that jetliners linked up to in the 1950s. Jet engines made it possible to fly at much higher altitudes, comfortably above the weather. A third trend was the development of larger aircraft, which were more cost efficient because of economies of scale. The DC-3 was extended into four-engine aircraft designs in the late 1940s, accommodating more passengers. Because jet engines were more powerful, they enabled the building of even larger aircraft, first the Boeing 707 and DC-8, and then the gigantic Boeing 747. This trend also allowed a decrease in cost, making flying ever cheaper. A fourth development was the lengthening of air routes. Long-distance markets (for example, across the Atlantic Ocean) gradually emerged in the 1930s with flying boats. In the 1940s, four-engine aircraft crossed the Atlantic with stops in Greenland, Iceland and Ireland. With the DC-7 (1953) non-stop transatlantic flights could be made. Jetliners linked up with the emergence of long-distance markets. These long-distance routes were the first market niches for civil jetliners (1958), because here they had the greatest benefit compared to piston engines. As jetliners became more fuel efficient, they further increased the travel distance and lower costs boosted the long-distance market, making the world a smaller place.

In the transition, external landscape developments also played a role. The Second World War was crucial for the emergence of jet engines. It allowed the 'hidden novelty' to acquire a supporting network and funding from the military. The Korean War was important, because it accelerated the American conversion to jet fighters after a confrontation with a jet-powered MIG-15. During the Cold War, strategic bombers acquired a central position in military strategies, especially the jet-powered B-52 bomber.

In sum, I conclude that the multi-level perspective has a good match with the case study, and therefore is robust.

Co-Evolution of Multiple Technologies

The system innovation in aviation took place because of interactions between multiple technologies. One pattern, already mentioned in Chapter 3, was the importance of incremental innovations alongside radical innovations. During the transition, many incremental innovations were made with regard to the main components of the jet engine. With regard to the compressor there were two competing designs: axial and centrifugal compressors. In the early days, centrifugal compressors were easier to build and operate, but bulkier and with a larger diameter; axial compressors had vulnerable blades, which often broke down. As stronger metal alloys were developed, the performance of axial compressors gradually exceeded that of centrifugal compressors. Blade design benefited from insights in aerodynamic science. Variable stators and the twin-spool concept were two important design innovations that helped to overcome the problem of compressor stall in axial compressors. Another innovation was the development of fans, which helped to increase the fuel efficiency and thrust of jet engines. With regard to the combustion chamber, many incremental innovations were made to its shape and the use of new heat-resistant alloys, which allowed higher temperatures. The functioning of the turbine also improved with heat-resistant alloys, as well as with better blade design.

A second pattern was cross-sectoral technology flows. Technical innovations were not limited to the aviation sector. The chemical and petroleum industries were important because they developed cheaper and better fuels. The commercial aircraft industry also borrowed technologies from the electronics industry, which since 1940 has provided a stream of crucial innovations ranging from radar to airline reservations and navigational computers. The metallurgical and materials industries were crucial, because they provided a wide range of new alloys and composite materials.

A third pattern was the importance of material innovations. In the late 1920s and early 1930s, duralumin was crucial for the shift from wooden to metal aeroplanes. In the 1940s and 1950s, new heat-resistant alloys were crucial for performance improvements in combustion chambers and turbines. New airfoils were important to improve the smoothness and aerodynamic efficiency of aircraft in the 1940s and 1950s. Titanium alloys were important in the 1960s to build fans which could stand great centrifugal stress.

A fourth pattern was complementarity between technologies. The proper functioning of jetliners depended on the development of complementary technologies. Swept-back wings were developed to overcome the compressibility burble on wing tips at high speeds. Smoother airfoils were developed to reduce drag. To overcome instability at low speeds during landing and take-off, innovations were made in wing flaps, tail and rudder configurations. The use of jet aircraft in aviation regimes also required further technical improvements, for example, in runways and air traffic control (radio, radar, computer). With the introduction of jetliners in civil aviation in 1958 the difference in speeds between planes increased. This made the job for air traffic controllers more difficult. Furthermore, the air space became more crowded as the number of aircraft increased. A safe introduction of jetliners and handling of more aircraft was possible because of the wider use of radar in ATC systems, linked to computers. Another example of complementary technologies was refuelling in the air and jet bombers. The development of jet bombers was slow in the late 1940s, because jet engines used almost twice as much fuel as piston engines. Refuelling in the air was a major stimulus for jet bombers, because it meant that the high fuel consumption became less important.

A fifth pattern was (unexpected) cross-technical linkages, which positively influenced each other (not strictly in a technical sense). In the late 1930s, radar played a crucial role in changing the performance specifications of interceptor fighters for British air defence. The implementation of radar along British coastlines changed the user preferences regarding interceptor fighters, leading to a need for aircraft with a powerful thrust, which could climb to high altitudes rapidly when the enemy was approaching. Thrust became more important than fuel efficiency. This shift in performance criteria provided opportunities for jet engines.

A sixth pattern was that multiple technologies competed with one another – not just (improved) piston and jet engines, but also turboprop engines. In the late 1940s and 1950s the turboprop also made inroads into civil aviation, resulting in a more complex substitution process. In the long-distance market the piston engine was rapidly replaced by the jet engine. In the medium-range market, the piston engine was gradually replaced by the turboprop. In the late 1960s the jet engine entered the medium-range markets and partially replaced the turboprop. On short distances, however, the turboprop continued to be used.

A seventh pattern was that gas turbines and piston engines not only had a competitive relationship. In the 1920s and 1930s the relationship was more symbiotic, as the gas turbine was used as an auxiliary add-on in piston engine aircraft as part of the turbo-supercharger.

Diffusion as a Trajectory of Niche Accumulation

Jet engines and jet aircraft diffused because they were used in subsequent application domains. The jet engine emerged and was initially developed in the military domain. Three niches were important in this domain: (i) fighters, (ii) bombers and (iii) tankers. Civil aviation was another major application domain where jetliners were first introduced in the long-distance niche in 1958. As the performance of jet engines increased, jetliners were both scaled down (Boeing 727) and scaled up (Boeing 747). Figure 6.21

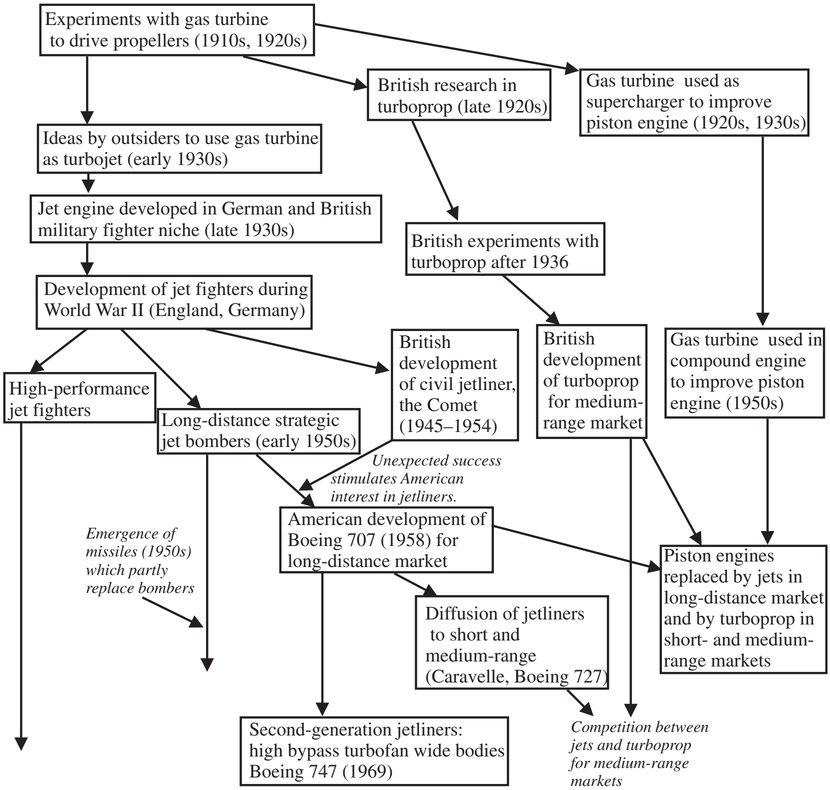


Figure 6.21 Diffusion of jet engines and jet aircraft as trajectory of niche accumulation

gives an indication of the trajectory of niche accumulation of jet engines and jet aircraft. From this figure it is clear that there was not one moment of breakthrough in the transition. Instead, the transition occurred as a sequence of larger and smaller steps.

Phases in the Innovation Journey of Novelty

In Chapter 3, I distinguished four phases in system innovations. In the turbojet transition these four phases did indeed occur, and therefore I conclude that they are corroborated. There is one aspect, however, which does not match with the predicted pattern, namely the articulation of new functionalities. The new functionality (flying for the masses) was not articulated in the second phase, but in the fourth phase.

First phase: emergence of novelty in the context of existing regime and landscape

The gas turbine emerged in the regime of electricity generation, where the focus was on the rotational power of the drive shaft. Following this guiding principle, the first ideas about gas turbines in aircraft were to use it to drive a propeller (turboprop). However, these ideas were not implemented in practice at that time. In the 1920s the gas turbine was used as an auxiliary device in turbo-superchargers to improve the functioning of piston engines at high altitudes. In the early 1930s, academic outsiders developed the ideas to the use the thrust of exhaust gases from the gas turbine. The idea for a turbojet was ignored by the existing aeronautic community.

Second phase: technical specialisation in market niches; exploration of new functionalities

In England, Whittle and his friends set up a small company dedicated to the development of the turbojet, and in Germany the pioneers found some interest from established aero-engine companies (Heinkel Aircraft, Junkers). These pioneers were uncertain about the functionality of jet aircraft. Whittle foresaw a small, fast, high-altitude transatlantic mail plane, while von Ohain envisaged military fighters.

Under the threat of the Second World War, the military interest in the turbojet increased. There was a proliferation of turbojet projects, which also involved established aero-engine companies. A dedicated community emerged to improve the turbojet. A market niche was found for the jet engine: the interceptor fighter.

Third phase: breakthrough, wide diffusion and competition with the established regime

After the war, there was a shift in the perception of the aeronautic community. The turbojet was perceived as having a future for high-performance applications, for example, fighters. Because the jet engine used a lot of fuel, it was thought unsuited for civil aircraft and there were doubts about using the jet engine for bombers. The introduction of jet engines in fighters was accompanied by wider changes in the aircraft, for example, swept-back wings and smoother airfoils. The thrust and fuel efficiency of jet engines was gradually improved. For bombers and civil aircraft there was a period of uncertainty about the engine of the future. Three designs competed: improved (compound) piston engines, turboprops and turbojets. In the early 1950s, jet engines came to be used more widely in strategic bombers, helped by refuelling in the air. In civil aviation there was also a period of uncertainty (1948–55). Britain introduced both turboprop and jet engines in civil aircraft in the early 1950s. Britain's jetliners (the Comet) enjoyed success and popularity, but the trajectory was halted by accidents in 1954. The British success triggered interest in the American aviation community. Pan American was the first airline company to show serious interest in buying jetliners. Boeing and Douglas designed jetliners in the mid-1950s in an innovation race. In 1958 the first jetliner (Boeing 707) entered American civil aviation on long-distance routes.

Economic perspectives emphasise that the diffusion process was driven by price/performance improvements. Incremental innovations raised the performance of jet engines. The thrust of jet engines increased almost 30-fold between 1937 and 1954. The protype jet engines by Whittle (1938) and von Ohain (1937) had a thrust of 480 and 551 lb, respectively. GE's J-79 (1954) delivered about 15,000 lb with an afterburner. The innovation of fanjets and larger BPRs increased the fuel efficiency and reduced fuel costs, stimulating the diffusion into civil aviation.

Socio-technical perspectives emphasise that diffusion took place because an increasing number of heterogeneous elements were linked together. This involved all kinds of technologies, but also aerodynamic science, government subsidies to research institutes and R&D programmes, military investment strategies, expansion of civil aviation markets and shifts in the symbolic perception of flying.

Sociological perspectives emphasise the importance of mechanisms in diffusion processes, for example, strategic games and changing perceptions. The case study has shown repeatedly how the perception of jet engines changed as practical experience was gained. In the 1930s the perception had hardened in the aviation community that gas turbines had no practical use, but this changed when jet aircraft were used during the Second World War. At that time they were seen only as useful for fighter aircraft, but the development of refuelling in the air and more fuel-efficient engines also made them applicable for bombers. The jet-powered B-52 bomber showed the aeronautic community that jet engines could power very large aircraft, but American civil airline companies still saw the jet engine as a useless gas-guzzler that would only cost money. It was only after some years of practical experience that this perception also changed. The jetliner could save costs on several dimensions, which were not foreseen (for example, maintenance, cabin crew, yearly profitability). The case study also showed the importance of strategic games and how they may speed up the transition process. The American civil aviation community was hesitant about jetliners, and did not invest in their development. Britain's companies were willing to take the risk, because they lagged behind America and because they were sponsored by the government. Britain's work on the jetliner greatly accelerated the American trajectory. American airline companies held each other in a 'cartel of fear': no one wanted to take the first step, because of the uncertainties and risks involved. However, Pan American's order for jet aircraft broke the inertia: because other American airline companies felt that they could not be left behind, they also ordered jets (domino effect). American aircraft manufacturers were also involved in strategic games. Boeing decided to build a jet tanker for the military, which it would also try to sell as a jetliner to the airline companies. The competitor, Douglas Aircraft, knew of Boeing's plans, but adopted a wait-and-see attitude. Thus Boeing gained a development advantage and was able to secure more orders for jetliners than Douglas, thereby overtaking Douglas's dominant position in civil aircraft manufacturing.

Fourth phase: gradual replacement of established regime, transformations and wider impacts

Although there were initial fears about high fuel costs, jetliners proved to perform cost efficiently (for example, lower maintenance costs, economies of scale). In the long-distance market, piston engine aircraft were rapidly replaced. In the short- and medium-range markets the turboprop replaced the piston engine. As jetliners were scaled down (Boeing 727) they entered medium-range market niches and replaced turboprops. Aircraft were also scaled up. The Boeing 747 (1969) introduced a new technical form, the wide body, which allowed a 30 per cent drop in tariffs. As airline companies searched for new user groups to fill their planes, they opened up a new market. Flying for the masses became a new functionality, which required changes in the entire socio-technical regime, for example, longer runways, larger terminals, lower tariffs, adaptations in air traffic control (radar and computers) and new skills for pilots. Jet aircraft also had wider impacts, in particular the expansion of global tourism: as tariffs dropped, more people flew to far-away holiday destinations.

Fit–Stretch Pattern in the Co-evolution of Form and Function

In a fit–stretch table I have summarised the co-evolution of form and function during the turbojet transition (see Table 6.2).

Table 6.2 Fit–stretch in the co-evolution of form and function in the turbojet and jetliner transition

Technical form / Use environment	Fit		Stretch
Fit	Functionality of gas turbine interpreted in terms of shaft rotations to power a propeller (1910s, 1920s).	Gas turbine used within piston engine aircraft as auxiliary device (supercharger) (1920s).	
		Aerodynamic scientists developed ideas to use the gas turbine for producing thrust (late 1920s, early 1930s)	During WWII, jet engine was developed in the interceptor fighter niche. Jet engine triggered changes in fighter aircraft, e.g. swept-back wings, smooth airfoils. Gradual use of jet engines in bombers (1950s)
			Jet engines gradually applied in civilian aircraft (late 1950s)
Stretch			Wider adaptations in socio-technical regime, e.g. runways, air traffic control, pilot skills, maintenance procedures, airfares, user groups. Expansion of world wide tourism; flying for the masses

7. Conclusions and discussion

7.1 INTRODUCTION

The three research questions that have constituted the topic of this book are:

1. How do system innovations and technological transitions come about?
2. Are there patterns in system innovations and technological transitions?
3. Are there particular mechanisms in system innovations and technological transitions?

Conceptual propositions about questions 1 and 2 were advanced in Chapter 3. This chapter begins by drawing conclusions about these conceptual propositions (Section 7.2). If the propositions have a good match with the empirical material in the case studies, they are robust. If there are deviations from the propositions, I shall advance refinements or additions.

The multi-level perspective provides the main answer to the first research question. In Section 7.3 I shall refine this answer by distinguishing three different routes in system innovations. This refinement is based partly on a further reflection about the three case studies, and partly on more detailed distinctions in niche-regime interaction. Section 7.4 answers the third research question about mechanisms. The case studies showed that all kinds of mechanisms were important in system innovations and technological transitions. Complementing the case studies with findings from the literature, I have organised these mechanisms with regard to phases in transition processes. The chapter ends with some reflections on the kind of knowledge claims and generalisability (Section 7.5).

7.2 CONCLUSIONS ABOUT CONCEPTUAL PROPOSITIONS

To answer the first research question, a multi-level perspective was described in Chapter 3. Several propositions were proposed about system innovations. First, novelties emerge in technological and/or market niches in the context of an existing regime and landscape. Niches are crucial for system innovation, because they provide the seeds for change. Second, the further success of a novelty is governed by processes within the niche and by developments at the level of the existing regime and the socio-technical landscape. Ongoing processes at the regime and landscape levels may create windows of opportunity for breakthrough. These ongoing processes may occur on any dimension of the socio-technical regime (technology, science, culture, policy, users preferences and markets). The following circumstances are often important to create windows of opportunity: (a) internal technical problems in the existing regime; (b) problems external to the system and policy measures to deal with them; (c) changing user preferences and emerging markets; (d) strategic and competitive games between firms; and (e) availability of complementary technologies. Two questions were proposed as proxies to test the usefulness of the multi-level perspective. How did niches emerge in the context of the existing regime? To which ongoing processes at the regime and landscape levels did the niche link up? To answer the second research question, the following patterns were proposed in Chapter 3: four phases in system innovation, co-evolution of multiple technologies, fit–stretch pattern in the co-evolution of form and function, and diffusion as a trajectory of niche accumulation. Below I shall draw conclusions about these six propositions.

Conclusion 1: Novelties Emerge in Niches

Novelties emerge in technological niches, shaped by the existing regime and landscape. The novelties often contribute to solving problems in the existing regime. The interpretation of the functionality of novelties often occurs with categories from the existing regime. Below I summarise the findings from the case studies, and conclude that this proposition is matched by the case studies.

Chapter 4 described how the first niche for steamboats was on inland waterways (for example, rivers, canals). On American inland waterways the steamboat was widely used in the 1810s and 1820s, because it linked up with the landscape process of westward settlements. In ports, the steamboat was used to improve the sailing ship regime, namely as steam tug to help manoeuvre large sailing ships. In the 1820s, steam machines were used incidentally as an auxiliary device (add-on) on oceanic sailing ships, to be used when there was no wind. In 1838 the British government created a new niche on the oceans when it decided to subsidise mail steamers to transport mail within the Empire, thus improving the coordination in trade and politics.

Chapter 5 described how automobiles emerged as technical novelties in the 1880s, and how they linked up with different niches in the 1890s. The early automobiles emerged as an add-on when the internal combustion engine, electric motor and batteries or steam engine were added to carriages or tricycles. The electric vehicle remained close to existing carriages in form and function. It was used in the urban luxury niche for promenading and tea parties. Between 1898 and 1902 the electric vehicle was also used in the taxi niche, but this was a failure. Gasoline automobiles (and steam automobiles to a lesser extent) built more upon bicycles, and were used in the racing and touring niches. In these niches the gasoline car developed a new form (low, wide, with the engine under the hood). These niches linked up with the existing cultural climate in which fun, entertainment and excitement were important values. It was only after 1905 that some users (for example, doctors, salesmen), began to use the automobile for utilitarian and business functions.

Chapter 6 described how first ideas to use gas turbines in aircraft focused on their rotational power, using it to drive propellers (turboprop engine). This focus on the rotational power came from the regime of electricity generation where the gas turbine was initially developed. The idea of the turboprop was studied seriously in the 1920s, but not tried out in practice. Gas turbines first entered aircraft as an add-on in the form of turbo-superchargers (1920s), thus improving the performance of piston engines in the thin air at high altitudes. In this arrangement, the focus was still on the rotational power of gas turbines, but this niche generated much useful knowledge for the later turbojet. It was not until the early 1930s that some academic pioneers shifted their focus in the study of the gas turbine, to look at the thrust of exhaust gases. This led to ideas about the jet engine. However, these pioneers had great difficulty in attracting the interest of actors in the established aeronautic community. Because the promise of the jet engine was not immediately recognised, it remained a 'hidden novelty' in the early 1930s. Only in the context of the approach of the Second World War did military actors develop more interest in high-performance engines, and the jet engine was gradually developed for the interceptor fighter niche.

Conclusion 2: Diffusion and Breakthrough Depend on External Circumstances

Diffusion and breakthrough of new technologies occur as the outcome of linkages between developments at multiple levels. To counter the bias towards new technologies, I proposed that more explicit attention should be paid to ongoing processes in socio-technical regimes. The existing regime is not only a barrier for radical novelties; ongoing processes in regimes can also create opportunities for them. Novelties can link up with the established technology as an auxiliary device (as add-on), with new regulations, cultural debates, newly emerging markets, scientific insights, strategic games in industrial networks and so on. Below I summarise the findings from the case studies, and conclude that this proposition is matched by the case studies.

Chapter 4 described how the landscape development of American westward settlements, together with a lack of good roads, provided a perfect context for the wide diffusion of steamboats on inland waterways in the 1820s and 1830s. The Irish potato famine (1845–48), the European political revolutions of 1848, and the Californian gold rush (1848) formed the background for emigration to America in the late 1840s, forming the first commercial market niche for oceanic steamships. The physical landscape change of the Suez Canal (1869) created a major market niche for steamers in freight transport, because sailing ships could not use the Canal and had to sail around the Cape. In the shipping regime there were several problems to which the steamship linked up: (a) limited regularity and predictability, (b) limited speed and (c) limited long-distance coordination in trade and administration. In the early nineteenth century a new social group emerged in the shipping regime: professional ship owners. Steamships linked up with this group in the sense that they began experimenting with line services in the 1840s. Another ongoing process was that of the lengthening of ships. As wooden ships became longer in the 1850s, problems of longitudinal strength were experienced. To deal with this problem, iron was gradually introduced in shipbuilding. The shift to iron thus linked up with the trend towards longer ships.

In Chapter 5, I showed that the emergence of automobiles linked up with many processes in the urban transport regime. Because automobiles emerged in a society in flux, they also linked up with many wider landscape developments. Gasoline automobile linked up with new user preferences in the transport regime: individual and flexible mobility. These preferences had been expressed in interaction with the bicycle. The application domains of racing and touring were opened up by the bicycle, and then the automobile linked up with them in the 1890s and early twentieth century. Suburbanisation was a trend which started with the horse tram and was strengthened by the electric tram. The automobile linked up with this trend and greatly reinforced it. The automobile could also build upon streets with smoother surfaces (for example, concrete, asphalt), which were initially constructed for bicycles and street sweepers in the context of the Parkway and the City Beautiful movements. The Good Roads lobby, which was initially started by bicycle clubs, literally prepared the way for cars. In the 1890s a symbolic shift occurred as streets came to be seen more as transport arteries. The automobile could link up with this trend, and further reinforced it, leading to a new urban transport system in the 1910s. Furthermore, the automobile could link up with a range of landscape developments. American cities grew rapidly because of urbanisation and immigration. This rapid growth led to the proliferation of slums, and social and health problems. Public health was seen as an increasingly important issue, giving rise to the so-called hygiene movement. Horses were under pressure because they linked up negatively with this issue. Bicycles and automobiles, on the other hand, were seen as positive ways to experience nature and inhale fresh air. One answer to urban social and health problems was to live in detached houses on separate plots of land, that is, suburban living. Many middle-class American families perceived suburban homes as a haven from the hurly-burly of the city. As real wages rose, a new middle class emerged with more money and more leisure time, preferably enjoyed in the form of entertainment. A new popular culture emerged, focused on entertainment, excitement, adventure and outdoor activities. In this context, the niches of car racing and touring in the countryside could become very popular. Because fresh air and light were seen as good for health, cities began to develop suburban park systems. The newly constructed parks and parkways provided suitable arenas for the early cars.

In Chapter 6, I showed how the jet engine linked up with several ongoing processes in the aviation regime. One such was the development of radar, and its implementation along the British coastline in the late 1930s. This led to new performance specifications for interceptor fighters. Fighters could stay on the ground until radar signalled the approaching enemy and the imperative need to become airborne. To gain height quickly, these aircraft needed powerful engines. Turbojets could link with this need. Another ongoing process in the aviation regime was the airframe revolution in the 1930s, which resulted in stronger and more aerodynamic aircraft. The improvements caused some scientists to speculate that future aircraft could approach the speed of sound. Because this would lead to great problems with the propeller (compressibility burble), these scientists arrived at a 'presumptive anomaly', namely that such high speeds were unattainable with propellers. Hence, these scientists sought new propulsion systems, which led some of them to the turbojet engine. The turbojet also linked up with the trend towards higher speeds. Since the 1930s, aircraft have flown ever faster. But propeller-powered aircraft suffered from the high-speed problem, caused by the compressibility burble around propeller tips. This problem was overcome with the jet engine, thus allowing the continuation of the trend towards ever-higher speeds. Jetliners also linked up with the trend of flying at greater altitude, above the weather to reduce turbulence and airsickness. In the late 1930s, four-engine aircraft with pressurised cabins flew higher, but they also experienced the high-altitude problem (less power for piston engines and less grip for propellers). It was to this conflict that jetliners linked up in the 1950s, flying comfortably above the weather with full power. Jetliners also linked up with the trend towards larger aircraft and the exploration of long-distance routes. In the transition, external landscape developments also played a role, mainly in the form of wars. The Second World War was crucial for the emergence of turbojet aircraft, because R&D budgets were raised and new performance criteria specified (linked to radar). The Korean War (1950–1954) boosted a large-scale American conversion towards jet-powered fighter aircraft. And the Cold War accelerated the development of jet bombers for strategic long-range bombing.

Conclusion 3: Four Phases in the Innovation Journey of Novelty

In system innovations I distinguished four phases:

1. Emergence of novelty in the context of existing regime and landscape (see above); interpretation of functionality with categories from existing regime.
2. Technical specialisation in market niches; exploration of new functionalities.
3. Wide diffusion, breakthrough of new technology and competition with the established regime. The diffusion process can be understood from different disciplinary perspectives, which highlight different aspects. The multi-level perspective emphasised that diffusion depends on external circumstances, which create windows of opportunity (see above). But diffusion is also stimulated by internal drivers. Economic approaches argue that diffusion is driven by declining costs and improved performance, which enhance its competitiveness in markets. Increasing returns to adoption play a role here. In socio-technical approaches diffusion is conceptualised as a process of linking heterogeneous elements together, for example, products, markets, infrastructure, regulations and subsidies. The diffusion of technologies is possible because more elements are linked together. The increasing number of linkages leads to 'momentum', irreversibility, mutual dependencies, lock-in and path dependence. Sociological approaches emphasise the importance of mechanisms in diffusion processes, for example, strategic games, the innovation race and social struggle. I shall come back to social mechanisms in Section 7.4.
4. Gradual replacement of established regime, transformations and wider impacts on society. This phase usually takes place gradually. One reason is that incremental innovations improve the cost/performance ratio of the new technology in a gradual way. Second, societal domains can consist of many market niches, with different selection criteria. It takes time to subsequently conquer all these markets. A third reason is that the creation of wider dimensions of the socio-technical regime takes time, for example, new infrastructures, user practices, policies and

organisations. Fourth, the capital intensity of existing technologies may have a delaying effect. If firms have sunk investments in complex production systems, they will not shift quickly to other technologies. Likewise, users will not buy expensive new technologies until investments in the old ones have been written off.

In each of the empirical case studies I concluded that these four phases did occur. Hence, the propositions are matched by the empirical case studies. However there are two small deviations to the general pattern. The first is that in Chapter 6 (turbojet transition) new functionalities did not emerge in the second phase, but in the fourth phase. I shall return to this in the next section. The second deviation is that with regard to replacement I found two additional complexities in the case studies.

First, old technologies may continue to exist for a long time in particular market niches, even after the new technology has become dominant. Replacement may thus be partial. This persistence of the old technology may help to reduce social unrest, since actors associated with the old regime do not immediately lose their job. Sailing ships were used in bulk freight transport well into the twentieth century (Chapter 4). Horses were used in freight transport until the Second World War (Chapter 5). Piston engine aircraft were used on short-range routes well into the 1970s, for example, as feeder routes to large airports (Chapter 6).

Second, old technologies may transfer to other market niches when they are replaced in mainstream markets. When sailing ships were replaced in the 1870s and 1880s in mainstream freight markets, they moved to bulk markets. These happened to be growing at the time, because industrialisation processes in European countries increased the demand for raw materials (Chapter 4). When steamships were replaced by jet aircraft in the 1950s in oceanic passenger transport, they moved to the cruise market which was just emerging. When gaslight was replaced by electric light, actors from the gas regime moved to the heating market (Schot, 1998).

Conclusion 4: Fit–Stretch Pattern in the Co-evolution of Form and Function

A particular cross-section of system innovations is the co-evolution of (technical) form and function. In Chapter 3 (Section 3.4) it was argued that this process follows a fit–stretch pattern. Emerging new technologies have a close fit with the existing regime, both in a technical sense and with regard to the interpretation of functionalities (phase 1). As producers and users build up experience with the new technology, new technical forms and new functionalities are explored (phase 2). As new technologies break through and replace the incumbent technology, the regime is stretched in the direction of new forms and functions (phases 3 and 4). In my case studies I found that it was easy to map the co-evolution of form and function in fit–stretch tables. Hence, I conclude that the co-evolution of form and function indeed follows a fit–stretch pattern.

A refinement needs to be made, however, with regard to the emergence of new functionalities. The case studies in Chapters 4 and 5 followed the predicted pattern, in which new functionalities emerged in the second phase of the transition. In the steamship transition the new functionality of line services was created in the late 1840s, long before the breakthrough in mainstream markets (1870s). In the automobile transition individual and flexible transport were new functionalities, which were widely enjoyed in the application domains of racing and touring – domains to which the gasoline car linked up. In the turbojet transition, however, new functionalities emerged late in the transition process. Initially the jet engine was mainly a technical replacement on the level of components, offering higher performance on (some of the) *existing* performance dimensions (speed, thrust, scale, range). The new component first changed the form of the aircraft (for example, swept-back wings, new airfoils). Only later in the process did new functionalities emerge which were embedded in the socio-technical regime: flying for the masses and worldwide tourism. These new functionalities had to do with scale. This means that there are two routes for the emergence of new functionalities (Figure 7.1): the first is through the interaction of users with the new technology early in the transition process; and the second is that a new technology starts as technical revolution, and new

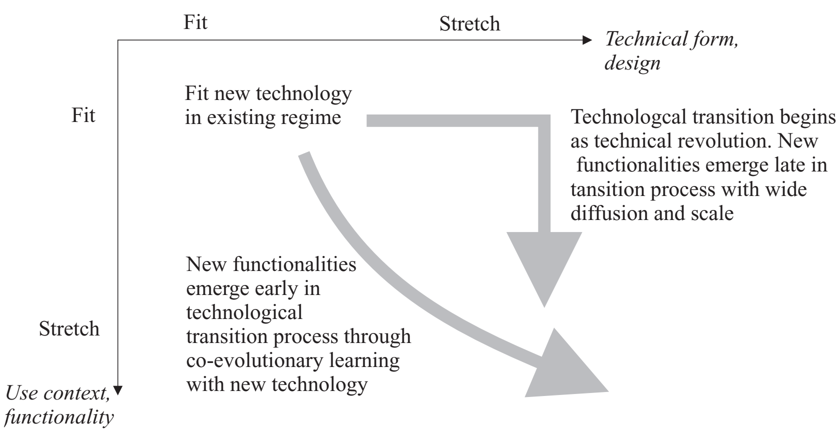


Figure 7.1 Two fit–stretch patterns in the co-evolution of form and function

functionalities are articulated late in the TT process. This is a refinement of the initial fit–stretch pattern, described in Chapter 3.

Conclusion 5: Co-evolution of Multiple Technologies

Transition processes are not about the emergence and breakthrough of one technology, but about linkages and interactions between multiple technologies. There is co-evolution between multiple technologies. In Chapter 3 (Section 3.4) I proposed several kinds of interaction:

- Interlocking and alignment between multiple technologies during diffusion. In Chapter 4, for instance, I showed how the diffusion of steamships depended on a combination of 'major' innovations: screw propulsion, iron hulls, compound steam engines.
- The importance of incremental innovations to increase the performance of a radical technology and stimulate its diffusion. The diffusion of steamships was greatly stimulated by new boiler designs, lubricants, anti-fouling paints and an adapted compass (Chapter 4). In my other case studies I also showed that incremental innovations are crucial for wide diffusion.
- Complementary technologies are important, that is, technologies which are necessary for the functioning of other technologies. In Chapter 4, I showed how the surface condenser was crucial for the use of the compound engine on oceanic steamships. In Chapter 6, I showed how the development of swept-back wings was crucial to enable jet aircraft to deal with turbulence and shock waves at higher speeds. I also showed how refuelling in the air was crucial for the diffusion of jet engines in bombers in the early 1950s. Also, radar had an indirect influence on turbojet development at the start of the Second World War, because it changed the Army's performance specifications for interceptor fighters. Radar also allowed the safe implementation of jetliners in civil aviation, because it improved air traffic control in the late 1950s.

Because these interactions had a good match with the case studies, I conclude that co-evolution of technologies is important in technological transitions.

On the basis of my case studies, I found the following additional kinds of interaction:

- Cross-sectoral technology flows were important. Technological developments in other domains may have important influences on system innovation. Innovations in steel were important to make stronger boilers and ship hulls (Chapter 4). The international telegraph influenced the steamship regime, because it further facilitated centralised fleet management. Innovations in gasoline, concrete, machine tools and fast-drying paints were important for the mass production and diffusion of automobiles (Chapter 5). Innovations in fuel, metal alloys and electronics were important for jet aircraft in aviation (Chapter 6).

- In all three case studies, transitions were accompanied by material innovations (for example, steel and anti-fouling coatings in steamships; fast-drying paints and float glass in automobiles; and duralumin, heat-resistant metal alloys and new airfoils in jet aircraft). Perhaps this is a typical finding for the transport domain, where strength and smoothness are important technical characteristics of the artefacts.
- In all my case studies I found the mechanisms of technical add-on and hybridisation. Novelties linked up with existing technologies as an auxiliary add-on to improve its functioning. Thus, old and new technologies did not immediately compete head on, but formed some sort of symbiosis (see also Pistorius and Utterback, 1997). Small initial mutations in existing technologies may over time result in new technical trajectories and new technical forms. In my case studies I found several examples of this pattern. The steam engine was used in sailing ships as an auxiliary device to address specific problems, for example, if there are no winds. In the 1840s and 1850s, hybrid ships were built using both sail and steam. Iron was first used in wooden ships as a complementary add-on, to strengthen the existing wooden constructions. As a second step, composite ships were built in the 1850s, consisting of an iron frame and wooden planking. The third step was an all-iron hull (Chapter 4). The first automobiles were no more than coaches or tricycles with an additional engine (Chapter 5). Gas turbines were used in propeller aircraft as supercharging devices to help them fly at high altitudes in thin air (Chapter 6). This pattern means that old and new technologies do not need to compete head on. Add-on and hybridisation are examples of symbiotic relationships between old and new technologies.
- Another pattern is the 'borrowing' of technical elements from competing technical trajectories. In 1911, gasoline vehicles borrowed the electric starter from the trajectory of electric vehicles. The improved gasoline car was able to invade the urban application domain. Another example is that early electric vehicles borrowed elements from several domains: (a) batteries and battery chargers came from

stationary electric applications; (b) electric motors came from the electric tram; and (c) the mechanical 'controller', a device which allowed easy gearing, was also borrowed from the electric tram. An example from Chapter 6 is that the turbojet borrowed from the turboprop the notion of combining gas turbines and propellers. This led to the turbofan in the 1960s. This pattern means that there can be cross-fertilisation between technical trajectories.

- Another pattern is that technologies can function as socio-technical stepping-stones, where they build upon each other in a socio-technical way. In the transition from horse-based transportation to automobiles, multiple technologies were involved. Horse-drawn carriages played a role, but also electric trams, bicycles, road surfaces and different kinds of automobiles. In Chapter 5, I concluded that automobiles were the final stage in a much longer transformation process. Gasoline cars could build upon all kinds of socio-technical changes, which had been expressed through the societal embedding of bicycles and electric trams. In particular, the bicycle was, in retrospect, an important stepping-stone towards the (gasoline) automobile. Bicycles led to the articulation of several new technical elements, for example, steel-tube frames, chain drive and differential gearing, ball bearings and air tyres. These elements were later used in early automobiles. The bicycle industry was also important in the articulation of techniques of quantity production utilising special machine tools, for example, sheet metal stamping and electric resistance welding. These new process technologies later became essential elements in the volume production of motor vehicles. With regard to user preferences, the bicycle led to an articulation of the preference for individual and flexible transport. The bicycle also opened new application domains, which later became linked to early automobiles: touring in the countryside and racing. With regard to social and infrastructural dimensions the bicycle gave rise to the creation of a Good Roads movement, which lobbied for streets with smoother surfaces (for example, asphalt), which inadvertently paved the way for automobiles. Following Sørensen (2002: 22), I suggested in Chapter 5 that the bicycle functioned as a catalyst in the automobile transition, opening up the transport regime and leading to many changes in the socio-technical regime. These changes created a fertile soil for the automobile. The implication of this catalyst effect is that even technologies that do not (eventually) win in terms of market shares, may be important for system innovations, because they introduce new elements into the socio-technical regime.

Conclusion 6: Diffusion Process as a Trajectory of Niche Accumulation

In Chapter 3 (Section 3.4) I argued that the diffusion process needs to be understood as a trajectory of niche accumulation, that is, new technologies are first used in particular niches or application domains, then in other niches, and eventually also in mainstream markets. This means that diffusion proceeds step by step. Because every step to another niche is like a jump, diffusion proceeds non-linearly in fits and starts. Each case study included a figure that represented these niche accumulation trajectories. Because this pattern had such a good match with the empirical case studies, I conclude that it is robust.

7.3 DIFFERENT ROUTES IN TRANSITIONS AND SYSTEM INNOVATIONS

The multi-level perspective gave a preliminary answer to the first research question, arguing that system innovations occur because of linkages between processes at multiple levels. This answer can be further refined by distinguishing different routes on the basis of differences in niche-regime interactions. This leads to three different transition routes: a) technological substitution with knock-on effects, b) wide transformation, and c) gradual transformation. These three routes can be further divided into sub-routes when taking into account other analytical dimensions. One such dimension is different kinds of technology, in particular the technical hierarchy and large technical systems. Other dimensions are the steepness of the learning curve and the occurrence of sudden landscape change, both of which influence the speed of transitions. The routes and sub-routes will be briefly illustrated with my own case studies and those of others.

First Transition Route: Technological Substitution with Knock-on Effects

In this route there is a particular niche–regime interaction. When the novelty emerges in a niche, the socio-technical regime is still relatively stable. There may be some minor problems in the regime, but these are not widely perceived as being very threatening to its survival. The shared perception is that these problems will be solved within the regime. The new technology is developed in particular niches, which have a low visibility at the regime level. In that sense, the novelties emerge 'below the surface', relatively hidden to actors at the regime level. These regime actors are not really looking for alternative technologies, because the existing regime is not experiencing major problems.

In the niches, learning and articulation processes take place, eventually resulting in a dominant design and substantial improvements in the new technology. While the new technology is being improved, the existing regime is still stable. But the improved technology may diffuse to other market niches. As the new technology enters mainstream markets, it begins to compete with the established technology. If the new technology replaces the old one, this is accompanied by knock-on effects, wider changes and adaptations, for example, in policy, infrastructure, user practices and industry structures. The technological substitution thus has knock-on effects throughout the socio-technical regime. Put simply, technological substitution in mainstream markets *precedes* wider co-evolution processes on other dimensions of the socio-technical regime. Hence, the route has a technology-push character. On the level of regimes this route is characterised by punctuation. Socio-technical regimes are relatively stable *until* the breakthrough and wide diffusion of new technologies. Adjustments lead to the formation of a new socio-technical regime, and then the dynamic shifts back to incremental change.

This basic technological substitution route can be further refined in three subroutes: (a) rapid breakthrough substitution; (b) gradual substitution; and (c) component substitution with innovation cascade. These subroutes will be briefly described below.

In the first subroute, breakthrough substitution, new technologies break through and replace existing technologies rapidly. This can be caused by steep learning curves in the new technology, which rapidly improve its performance and its competitiveness. It can also be caused by sudden changes at the landscape level, which rapidly change the selection environment in the regime. Or it can be a combination of both. Also in this route, the new technology emerges in a niche, with low visibility for regime actors. The work on novelties occurs mainly 'below the surface' in fringe niches. The dynamic can metaphorically be described as a 'peat moor fire'. As long as niche actors invest in learning processes and network building, the novelty may smoulder beneath the surface, but when a certain performance threshold is passed or when external circumstances change the selection environment, the new technology can suddenly break into mainstream markets and replace the existing technology. The rapid replacement may lead to the downfall of established firms and trigger Schumpeter's 'gales of destruction'. Although the breakthrough may be rapid, it was preceded by a long gestation period and is followed by wider adjustments. This breakthrough route is schematically represented in Figure 7.2.

An example is the transition from coal-based gas (city gas) to natural gas in the Netherlands (1959–70) (Correljé and Verbong, 2004). The transition to natural gas was a real system innovation with changes in the insti-

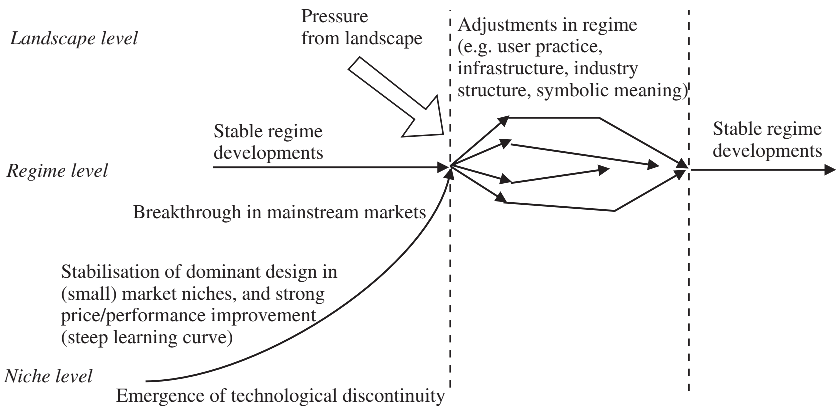


Figure 7.2 Breakthrough substitution route with knock-on effects

tutional framework of the gas industry, the distribution network and articulation of new markets, for example, central heating, cooking. Before 1959 there was already a small city gas network, based mainly on local infrastructures: 'long-distance' gas was offered by the Dutch steel works, Hoogovens, and Dutch State Mines (DSM) as a byproduct of their main activities. Thus there was an established gas regime. In 1959 a huge gas supply was discovered in Slochteren, which sent shock waves through the existing regime. In line with the political culture at the time, the Dutch government set up a broad consultation process to discuss how this natural gas should be exploited. Many actors were also invited from the existing regime (for example, Hoogovens, DSM, Exxon, Shell and its subsidiary NAM). Negotiations and strategic positioning resulted in a master plan for the rapid introduction of natural gas, involving agreements about markets, the institutional set-up and distribution infrastructure (Correljé and Verbong, 2004). It was decided that the gas would be sold not just to large industrial users, but also to citizens. This required the construction of a countrywide high-pressure transmission infrastructure that would link all local distribution systems to the Slochteren field. The Dutch government invested much money in this infrastructure. The institutional dimension was also developed, defining the roles and relations among the several actors involved. An important aspect was that losers were compensated, thus limiting their resistance: Hoogovens was compensated financially and DSM was given a role in the exploitation of the natural gas. Another institutional aspect was the setting of the price of natural gas (which was bound to the oil price) and the distribution of the revenues between state and private parties. The rapid diffusion was stimulated by pricing natural gas lower than city gas. In addition, the government set up a large information campaign to entice domestic customers to switch to gas. Meetings were organised, information leaflets were distributed and publicity appeared in the press and on the radio. Subsidies were given to consumers who had to adjust their cooking stoves and pans (because natural gas had a higher caloric content, the outflow speed of gas stoves should be lower than with city gas). The natural gas transition had many wider impacts: it was the deathblow for the Dutch mining industry; it boosted the state revenues and enabled the expansion of the welfare state; the Dutch government used natural gas as a policy tool to attract energy-intensive businesses (for example, chemical, oil companies, horticulture); and the widespread supply of gas enabled the upgrading of houses, with the introduction of central heating. In sum, the transition to natural gas followed a rapid breakthrough route, followed by wider adjustments. Other instances of the breakthrough substitution route can be found in Christensen (1997), who gives examples from the disk-drive, steel making, mechanical excavators and motorcycle industries.

The second subroute is gradual technological substitution. The new technology is developed in niches, where it can remain for a long time. The price/performance ratio of the new technology is improved incrementally, and the learning curve is less steep. Furthermore, landscape developments may change gradually, increasing the pressure on the regime slowly and stepwise. As a result of both developments, the competitive position of the new technology improves little by little. As the new technology conquers a greater market share, the old technology withers away, but may hold on to particular market niches for a long time. The replacement process may also proceed gradually, because adjustments in the wider socio-technical regime take time (for example, investments in infrastructure). This gradual substitution route is schematically represented in Figure 7.3.

An example of this route is the transition from sailing ships to steamships. The sailing ship regime was relatively stable. There were certain problems in the regime (for example, limited speed, regularity and long-distance coordination in colonial trade and government), but these were not widely perceived as threatening. Steamships plied the oceans in 1838 as subsidised mail steamers. Then in the late 1840s they came to be used in the passenger market, which rapidly grew because of emigration from Europe. The performance of steamships was gradually improved over time. The coal efficiency of steam engines improved, iron hulls allowed the building of larger ships which enjoyed economies of scale and screw propellers increased the transmission efficiency of power to motion. Following the opening of the Suez Canal, steamers also entered the freight market, which was enlarged and diversified because of gradual landscape developments, for example, industrialisation and the liberalisation of trade. The diffusion

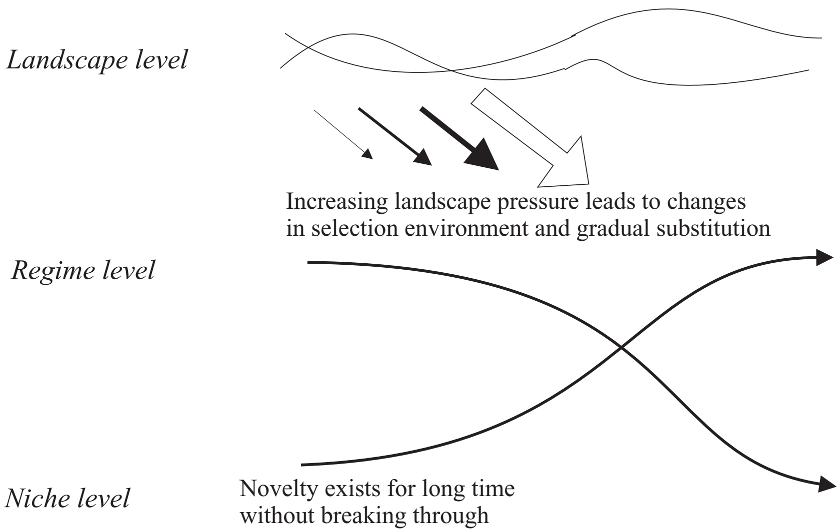


Figure 7.3 Gradual substitution route with knock-on effects

of steamers in the 1870s and 1880s was accompanied by all kinds of adaptations in the socio-technical regime, for example, deepening and enlarging of ports, new machines for loading and unloading, larger shipbuilding yards where new competencies and new machines were needed, and the institution of a worldwide coal infrastructure.

The third subroute is the bottom-up component innovation cascade. In the discussion of the literature on complex products and systems (see Chapter 2) I noted that the structure of complex technical configurations is that of a technical hierarchy, consisting of many components, devices and subsystems. Sometimes components are so drastically changed or improved that they require changes in other components or even a change in the entire product architecture. This idea forms the basis for this subroute. Substitution of a component may trigger wider innovation cascade dynamics upward in the entire artefact. Changes may thus cascade from lower to higher levels in the technical hierarchy. When such cascades go even further, they may influence the wider socio-technical system, thus contributing to a system innovation.

The transition from piston engine propeller aircraft to turbojets has a reasonably good match with this route. When turbojets emerged in the mid-1930s, the piston engine propeller regime was relatively stable. Although there were some minor problems (caused by altitude and high speed), the aviation community was convinced of the potential of piston engines and propellers to deal with them. But a sudden landscape development, the Second World War, created a particular niche, high-speed interceptor aircraft, in which the turbojet was adopted and further developed. After the war, the turbojet was further improved in the military domain, first in fighters, but by the late 1940s and early 1950s also in bombers. Ideas concerning the use of turbojets in civil aviation received much attention in the mid-1950s, after the relative success of Britain's Comet. In 1958 civil jetliners entered the long-range market niche, and subsequently diffused to other market niches. In terms of the bottom-up component innovation cascade, this began with a component substitution: the piston engine and propeller were replaced by a jet engine. This substitution subsequently triggered further changes in aircraft, for example, new wing designs (swept-back wings), new skin materials, and different wing flaps for landing and take-off. If change had remained limited to these components, it would have been a 'technical revolution'. But wider changes occurred in the socio-technical regime. As jetliners were used more widely in commercial aviation, longer and stronger runways were constructed, pilots required new skills, new user groups and markets were identified, leading to a new functionality (flying for the masses), new maintenance procedures were developed, helping to save costs, and new technologies were introduced in air traffic control systems (radar, computers). In sum, the transition had a technology-push character, which was driven by a component substitution.

Second Transition Route: Wide Transformation

In the second route the socio-technical regime opens up before new technologies emerge in niches, an important difference with the first route, where the regime was stable. This opening up can be the result of persistent internal problems, negative externalities and policy measures, changing user preferences, strategic and competitive games, and pressure from the landscape level (for example, new cultural values). As a result of these processes the regime becomes more fluid and opens up, creating windows of opportunity for new technologies. The problems in the regime provide a legitimisation for niche actors to work on innovations. To obtain funding, niche actors make promises about how innovations will contribute to solving the regime problems. Wider co-evolution processes in the socio-technical regime thus precede the emergence and breakthrough of new technologies.

Characteristic of this route is that the loosening up of the existing regime may create *multiple* windows of opportunity, leading to the exploration of multiple new technologies. New technologies may link up with the ongoing regime processes, and further contribute to the widening. This means that there are strong interactions between niches and regimes. When there are problems in the existing regime, regime actors are likely to be more willing to look for alternatives. Experiences and learning processes in niches may influence perceptions, expectations and strategies of actors in the regime. Regime actors may adopt hedging strategies, and begin experimenting with new technologies. As a result the fluidity in the regime increases, thus creating further opportunities for niches. These niche–regime interactions may lead to gradual transformation and further widening of the regime. There may be a prolonged period of experimentation with many novelties co-existing. Such a period of heating up and experimentation is usually followed by a 'cooling down' period, that is, a narrowing of the number of technical options. A particular technical option may come to be seen as 'universal', pushing other options out of the market. At the regime level there are four phases in this route: (a) opening up; (b) increasing technical variation, wide experimentation, co-existence of multiple options, uncertainty; (c) cooling down; and (d) stabilisation of the new socio-technical regime. The wide transformation route is schematically represented in Figure 7.4.

In this route there is no technology push. Instead, the socio-technical system is transformed on many dimensions simultaneously, for example, cultural values, user preferences, markets, policies, industry structures, infrastructures and technologies. There is a prolonged period of heating up and cooling down. System innovations take a long time in this route, because many elements are first decoupled and then linked together again.

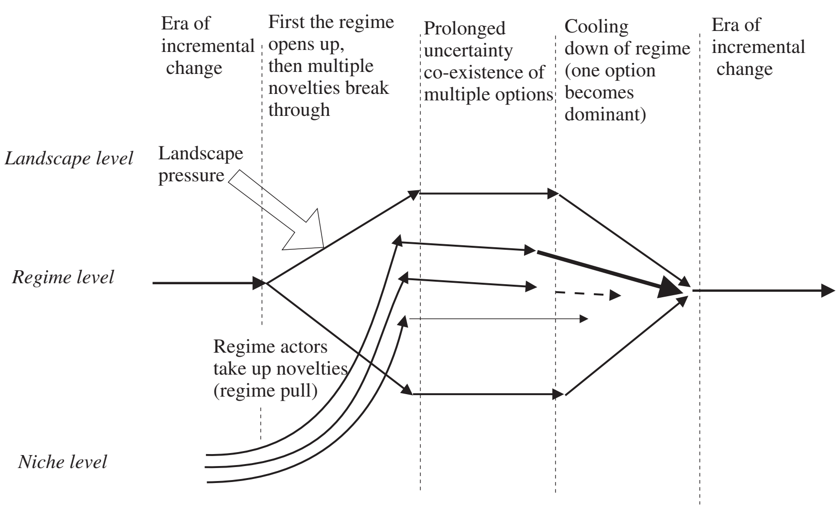


Figure 7.4 The wide transformation route

The transition from horse-drawn carriages to automobiles has a good match with this transition route. Chapter 5 clearly showed how the urban transportation regime was heating up in the late nineteenth century, because it was experiencing several problems: (a) pollution by horse excrement and concerns over hygiene; (b) congestion in the streets; (c) high cost for feeding, operation and maintenance of horses; and (d) accidents. These regime problems were made worse by landscape developments. The rising concern about public health at the end of the nineteenth century, led to debates about horse excrement on the street. Urbanisation at the end of the nineteenth century led to larger cities and longer travel distances. Coupled with congestion, horse-based transportation had increasing difficulty in meeting the demands for faster transport. Suburbanisation at the end of the nineteenth century was a social process in which the middle classes left the congested cities to live in the suburbs. This increased the distances they had to travel, and put further pressure on horse-based transportation. Because of these problems, the regime had been opening up since the 1860s and different technical innovations were introduced, for example, the horse-drawn omnibus and the horse tram. In the 1880s and 1890s, other innovations emerged, for example, steam and electric trams, bicycles, steam buses and steam, electric and gasoline automobiles. The 1890s saw a marked widening up of the urban transport regime, characterised by the co-existence of multiple transport options. But the widening up also involved wider changes in the socio-technical regime. Some of these social, political, institutional and cultural changes preceded technical changes, others developed in co-evolution with new technologies, in particular the bicycle and the electric tram. With regard to user preferences, the bicycle led to an articulation of a preference for individual and flexible transport. It also opened new application domains: touring (in the countryside) and racing. Suburbanisation was stimulated in the 1890s by the emergence of the electric tram, creating new mobility patterns (commuting to work, shopping in the city). The electric tram also stimulated a cultural change in the perception of the function of streets (from social meeting place to transport artery). Another cultural change was that city residents became more accustomed to faster vehicles, in particular the electric tram. Another change took place in public administration, where the authority of city government was expanded. This enabled city authorities to push through the implementation of roads with a smoother surface, against the wishes of the local residents. These smoother roads initially benefited cyclists and street sweepers, but later paved the way for the automobile. On the political dimension there were several reform movements that strived for a solution to urban problems (for example, slums, pollution, disease). They promoted suburban living and the creation of new public spaces (for example, parks, boulevards), which later provided arenas for the first use of automobiles. In sum, the socio-technical regime was already changing substantially *before* the emergence of automobiles.

So the first phase of the transition from horse-drawn carriages to automobiles was characterised by a widening up of the urban transport regime. In terms of technical transport options, the transition was a widening of transport technologies that co-existed between 1880 and 1910, for example, horse and electric trams, bicycles and automobiles. The electric tram regime became increasingly important in terms of total mobility, leading to a narrowing down of technical options. But although the electric tram acquired a prominent position in urban transport, it was not the final winner. In the 1920s the electric tram regime disintegrated, and the automobile regime became dominant. The automobile was thus the final stage in a wider and longer transformation process. Figure 7.5 schematically represents the transformation route for the automobile transition.

Third Transition Route: Gradual Reconfiguration in Large Technical Systems

Large technical systems are a particular kind of technology, consisting of interrelated technologies, often with infrastructural networks, which

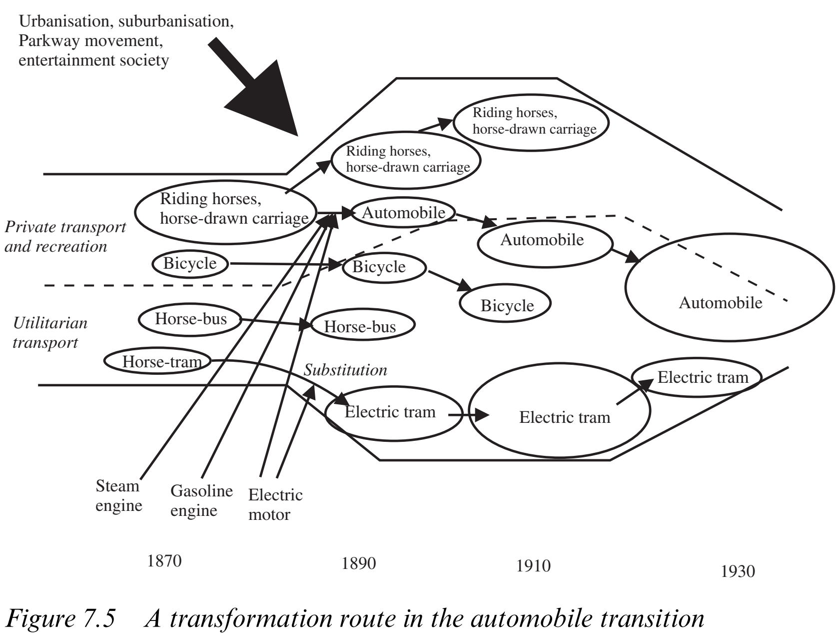


Figure 7.5 A transformation route in the automobile transition

stretch geographical areas (see also Chapter 2, Section 2.2). Examples of LTS are electricity systems, telephone networks, railroad systems and the Internet. LTS are likely to require a different transition route, namely gradual reconfiguration. Because of its systemic character, an LTS needs to maintain its integrity, in the sense of functional linkages between elements. In a gradual reconfiguration route the old and new technologies closely interact from the start, as add-on and hybridisation. A new innovation first links up with the old system as an add-on, playing a minor, auxiliary role. The roles of old and new technologies may change gradually during a reconfiguration process, due partly to external circumstances and partly to technological improvement of the new technology (see Figure 7.6).

In electricity production, for instance, gas turbines were first used in the small market niche of peak loads. Subsequently they were used as an auxiliary device to improve the performance of the steam turbine (combined cycle power stations), and then in the late 1980s the external circumstances changed (Winskel, 2002). Liberalisation in the electricity sector created more uncertainty and changed the criteria used for investment decisions, for example, quick returns on investment, shorter construction times and modular construction mode (which increased flexibility). These changes stimulated the construction of gas turbines at the expense of steam turbines. Another landscape change was the growing concern about the environment, something which also stimulated gas turbines, since gas was a cleaner fuel than coal, which was used in steam turbines. An increase in the discovery of natural gas supplies was another external development that stimulated gas turbines. Also, the performance of gas turbines improved over the years. As a result of these interacting processes, gas turbines became the main component in the combined cycle, while the steam turbine took on the role of an auxiliary device (Islas, 1997). Thus examples of the third route can be found, and it is a plausible option.

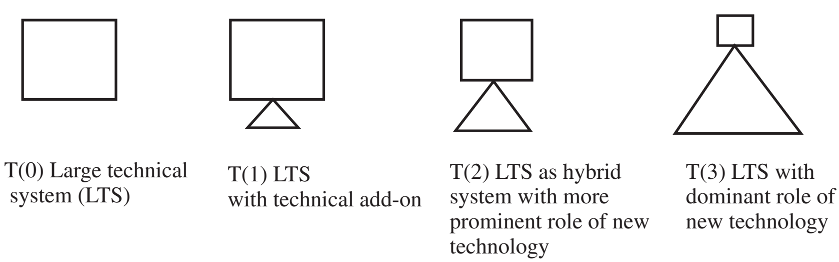


Figure 7.6 Reconfiguration dynamics in large technical systems

7.4 MECHANISMS AT PLAY IN THE UNFOLDING OF TRANSITIONS AND SYSTEM INNOVATIONS

This section addresses the third research question: are there particular mechanisms in system innovations and technological transitions? The multi-level perspective gives a general answer to how transitions come about. The explanation of the emergence of new regimes is that multiple developments gradually link up and reinforce one another. The explanation is thus located in the process, in the alignment and interlocking of different processes. The convergence and alignment of processes increase the chance of regime transformation; they create the 'right' circumstances. But the actual linkages between the processes need to be made by actors in their cognitions and activities. To give precise explanations of system innovations, the multi-level perspective needs to be filled in with more detailed accounts of the role of actors. Mechanisms are a way of introducing actors and their activities into the story, for example, strategic games and the sailing ship effect; they can speed up/slow down certain processes, or lead to changes in directions of developments.

In my case studies I encountered many mechanisms in system innovations. In this section I shall list a number of relevant ones, combining insights from the literature with findings from the case studies. The list is organised in terms of the four transition phases:

- first phase: the emergence of novelties in niches;
- second phase: the specialisation and take-off of new technologies;
- third phase: wide diffusion, breakthrough in mainstream markets, competition with the existing regime; and
- fourth phase: replacement of the existing regime.

For the fourth phase I have already proposed some mechanisms in Section 7.2 (refinements of the third conclusion).

First Phase: The Emergence of Novelties in Niches

1. Add-on and hybridisation (see also Section 7.2, refinements of conclusion 5). New technologies link up with existing technologies as auxiliary add-ons, subsequently leading to hybrid technical forms. Thus, old and new technologies do not compete head on, but initially have symbiotic relationships.
2. Spatial and geographical proliferation of technologies (also Schot, 1998). When a niche in one country dies, the technology may develop further in another country. The first steamboat experiments in the 1780s

occurred in Britain and France, followed in the 1890s by America. When the steamboat niche in Britain and France died out (because of the French revolution, and resistance by British canal companies) experiments continued in America, leading to a market niche in 1807. Automobiles were initially developed in Germany and France (in the 1880s). By 1895 the first automobiles were introduced in America. After 1900, America jumped ahead in car sales. The jet engine was first developed in Germany and Britain in the 1930s. America began working on jet engines in 1941.

3. Outsiders (for example, young people, small companies, academics) are important in the creation of radical novelty and niches. Members of the established technical community often have vested interests, or are blinded by established cognitive routines and search heuristics. Although outsiders are important in the generation novelties, many of them fail, because they lack the financial and organisational means to develop their ideas (see Olleros, 1986). One example is that ideas about jet engines were developed by academic outsiders (Whittle, von Ohain) who had difficulty in finding funding for their ideas, because the established aeronautic community was convinced about the potential of piston engines and propellers. Another example is that Britain, a small player in the international aircraft industry after the Second World War, used its expertise in jet engines to develop a civil jetliner to break the American dominance. A third example from Chapter 6 is that Boeing, a relative outsider in civil aircraft construction, expected their main market to shrink (bombers), so aimed to enter civil aviation with jetliners. An example from Chapter 5 is that Ford, a relative outsider in the automotive community in the early twentieth century, began to experiment with cheap, strong cars in 1903 while the established automobile community focused on luxury cars. After years of experimenting this led to the Model T in 1908.
4. Saturation of existing markets, sales crises and expectations of shrinking markets provide a stimulus for actors to diversify to other markets and technologies. The bicycle crisis of 1898 stimulated bicycle producers to diversify into automobile production. The automobiles sales crisis of 1907 signalled the saturation of the luxury market, and triggered a more utilitarian design trajectory, leading to the Model T. In the early 1950s, Boeing expected that the market for heavy bombers would shrink because of the emergence of cruise missiles. It therefore decided to develop jetliners for the civil market.
5. The importance of policy support in the creation of niches and keeping them alive. National governments gave tremendous financial support to airline companies in the 1920s and 1930s (via direct subsidies in

Europe and indirectly in America with lucrative mail subsidies). Governments also stimulated aviation via the sponsoring of research (aerodynamic, fuels, engines). Another example is that the first niche for oceanic steamships was created by British mail subsidies after 1838. A third example is that automobiles in the early twentieth century came to be seen by progressive politicians and urban reformers as a means to achieve the goal of suburbanisation. Hence, automobiles were regulated favourably. In the 1910s and 1920s, large sums were used to build roads and highways (at the expense of electric tramways).

6. The role of wide cultural visions and promises. When new technologies link up with wide cultural visions and values, this creates a legitimisation and protected (cultural) space to work on new technologies (see also van Lente, 1993; Chapter 5). Aircraft in the 1920s and 1930s enjoyed great popular support, because they were seen as means to a better world, the winged gospel. Another example is the gasoline car, which linked up with cultural values such as excitement, danger and adventure.

Second Phase: The Specialisation and Take-off of New Technologies

7. Technologies break out of their niche by linking up with the (sudden) growth in particular market niches. The growth of (new) markets may provide opportunities for new technologies to ride piggyback. In the early twentieth century the sales of gasoline cars ran ahead of electric vehicles and steam automobiles, by linking up with the strong growth in the market niche of touring. Oceanic steamships could take off commercially in the late 1840s when the passenger market to America grew quickly.
8. The emergence of specialised social groups can be instrumental in the diffusion of new technologies. Such groups will do their best to advance the technology, either in a technical sense or through political lobbying. Specialisation and professionalisation may be important processes in their emergence. An example from Chapter 4 is the emergence of professional ship owners in the late eighteenth and early nineteenth centuries. Because ship owners ran commercial shipping services, they had a vested interest in fast ships that can make more trips per year and earn more money. Ship owners were also the first group to use steamships for oceanic passenger transport. The emergence of highway engineers and their professionalisation within civil service organisations after 1910 was crucial for the later expansion of road building. An example from Chapter 5 is the emergence of highway engineers within civil service organisations after 1910, and their role in the expansion of road building.

9. Taking large steps in the design of new technologies (rapid upscaling) may be risky when their performance is not well established. To break the American dominance in aircraft building, the British government opted for a 'leapfrog' strategy, using their expertise in jet engines to develop a civil jetliner, the Comet, which entered commercial service in 1952. After two years of great popularity, the Comet suffered fatal accidents, which were related to metal fatigue around the square windows, a relatively unknown phenomenon at the time. An example from Chapter 4 is the steamship *Great Eastern* (1858), which represented a gigantic leap forward, being six times larger than any other ship. However, it experienced several problems: the coal consumption had been seriously underestimated, and the ship unexpectedly rolled excessively in bad weather.
10. Strategic games and innovation races can play a role in the acceleration of technical trajectories (domino and bandwagon effects). Companies or governments decide to sponsor certain technologies for fear of lagging behind. When Germany became an industrial power at the end of the nineteenth century, it entered into competition with Britain to develop the fastest and largest steamships. The innovation race speeded up the emergence of superliners in the early twentieth century. An example from Chapter 6 is that the British government wanted to break America's dominance in aircraft production after the Second World War, and stimulated the development of civil jetliners. The positive experience with the Comet changed perceptions in the aviation community and accelerated the development of jetliners. Another example from the same chapter is that Boeing and Douglas were involved in a strategic game with regard to civil jetliners in the 1950s: while Douglas chose a wait-and-see strategy, Boeing embarked upon the jetliner path more actively, fearing that its core business of heavy bombers would shrink in the future because of the emergence of strategic missiles. Once Boeing had embarked on serious negotiations with Pan America, Douglas also speeded up its jetliner activities. On the user side, there was a domino effect. Airline companies long hesitated to order jetliners, because they expected higher fuel costs, and as long as nobody ordered jetliners, they were not needed. The 'cartel of fear' led to inertia, which was broken when Pan Am ordered jets in 1954. Other airline companies quickly followed suit, because they feared being left behind.
11. Social struggles over technologies can lead to delays or accelerations. The transition in the unloading of grain ships from manual labour to grain elevators could have occurred by the late nineteenth century, because all the ingredients were present. However, it did not occur

then, because economic gains were deemed too little, and satisfaction with the established regime was high. In 1907, however, grain elevators became part of a wider power struggle in the Rotterdam port between labourers and the unloading companies. To show who was the boss, these companies quickly shifted to grain elevators, using it as a weapon in the struggle (van Driel and Schot, 2001). In the transition from electric trams to cars and buses in the 1930s, there was a political struggle between tram companies and city governments in which buses were used as a pawn – city governments adopted buses to put pressure on tram companies. Chapter 6 described how radar had been developed during the Second World War, and was used as a ground controlled approach (GCA) to help pilots to land safely. In civil aviation, however, landing was based on the instrument landing system (ILS), whereby pilots found their own way using radio technology. After the war there was a struggle over the introduction of radar in air traffic control and landing systems. Civil pilots resisted GCA, because it reduced their autonomy. ILS thus won the battle, which delayed the introduction of radar, and it was not until a major accident in 1956 that radar came to be used more widely in air traffic control.

12. Hypes, bandwagon effects and self-fulfilling prophecies can play a role in the take-off of new technologies. During the bicycle craze (1895–97) many middle-class consumers bought bicycles to tour in the countryside. The opening of the Suez Canal and the associated opening of the Indian market triggered the steamship mania (1869–74), leading to accelerations in the diffusion of steamships in freight markets. Internet companies and UMTS (third-generation mobile telephony) are recent examples that show how hypes influence investment strategies by firms. When expectations are hyped too much, they may be followed by a backlash. After the bicycle craze, the global bicycle market collapsed in 1898 (saturation and overproduction), leading to price dumping and a rapid spread of bicycles to the working classes. In 1874 the steamship mania ended, and sailing ships enjoyed a (brief) revival. The Internet and UMTS hypes were followed by a backlash, leaving telephone companies heavily in debt.

Third Phase: Wide Diffusion and Breakthrough, Competition with the Existing Regime

13. During the diffusion process a dominant design may be stretched to accommodate the requirements of different market niches. After the automobile had become the dominant land-transport mode in America, its design was stretched in the 1930s, leading to a great

variety of shapes and form, but using the same basic technology. After the turbojet was introduced in commercial aviation in 1958 (the Boeing 707), the design was adapted downwards to the smaller 727 (1964) and upwards to the wide-bodied 747 (1969).

14. The sailing ship effect refers to the pattern by which the existing technology is improved when it is challenged by an emerging technology. Sailing ships in the 1860s and 1870s were strongly improved when they were challenged by steamships. To increase their speed, more masts and sail were added to sailing ships, and the hulls were redesigned. To reduce labour costs, labour-saving machines were introduced, for example, to rig the sails. An example from Chapter 5 is that the tram industry made a major effort to modify its technology in the mid-1930s to meet competitive pressures from the automobile and the bus. In 1935 the PCC car, a high-speed tram car of high capacity for main routes was introduced. However, it was a failure and trams were gradually replaced by automobiles and motor buses in the following decades.
15. The flip side of the sailing ship effect is that actors in the established regime hold onto the existing technology for too long, continually making improvements. This may lead to complex ‘monsters’. More and more masts and sail were added to sailing ships in the late nineteenth century, which led to some very unstable ships. The *Thomas Lawson* (1902), for instance, was a clipper ship with seven masts; it was very fast, but also unstable, and it capsized in 1907.
16. ‘Missing the wave’ means that established companies fail to take advantage of the wave of newly emerging technology, because they hold on too long to the old technology (Bower and Christensen, 1995). In this way, incumbent companies may go under. Most established shipbuilders did not join the transition from wood and sail to iron and steam, which required new skills (for example, riveting and metal working) and new machine tools, but continued to build wooden ships. As a result, the centre of gravity in British shipbuilding moved north to the Clyde and the northeast of Britain. An example from Chapter 6 is that only one of the two dominant American aero-engine manufacturers, Pratt & Whitney, made the transition from piston engines to jet engines. Curtiss-Wright went out of the business, while General Electric stepped in as a newcomer.

7.5 DISCUSSION AND REFLECTION

What kind of theorising has been practised in this book? Nelson and Winter (1982) introduced the notion of ‘appreciative theory’ in opposition to mainstream economic theory, which strongly adheres to formal theorising. They argued that formal (modelling) approaches downplayed the complexity of empirical reality and the importance of qualitative elements. Therefore they proposed 'appreciative theory' as another mode of theory development, which 'appreciates' the complex and sometimes fuzzy nature of reality. Nelson characterised appreciative theory as follows:

> This kind of (economic) analysis, which stays relatively close to the data and is expressed mostly in words, but which nonetheless involves theorizing. (Nelson, 1994a: 48)

> While starting with the empirical subject matter, the accounts put forth by economists of the development of an industry, or the evolution of a technology, focus on certain variables and ignore others. Quite complex causal arguments often are presented as parts of these accounts, if generally in the form of stories. (Nelson, 1995: 50)

In appreciative theory the aim is to understand a phenomenon in its empirical complexity. That is what I have done with system innovation and technological transitions. I did not develop a formal model, but 'appreciated' how different perspectives can make different contributions to understanding system innovations.

The appreciative theorising in this book has some strengths and weaknesses. The first strength of the multi-level perspective, in the extended version developed here, is its scope and generality. The perspective is encompassing and able to combine contributions from sociological, economic and socio-technical theories. Another strength is that the perspective is able to accommodate empirical reality, even if this is complex (see the case studies). A weakness of appreciative theorising is the use of metaphors and loose concepts. To further develop a future research programme, some concepts might benefit from analytical refinement. One candidate for this is the notion of 'opening up' of regimes, leading to 'windows of opportunity'. Another is the notion of 'socio-technical landscape'. At present it functions somewhat like a residual category that includes a broad range of phenomena, for example, institutional, material and cultural. For empirical research this worked sufficiently as a productive heuristic, as the case studies showed. For analytical statements, however, the concept needs to be further refined. A problem for analysts who are interested in making models, might be the perspective's limited degree of simplicity. In fact the perspective is fairly complex, requiring attention to dynamics at multiple levels and dimensions.

What is the scope of the conclusions? Do the conclusions and conceptual perspective hold only for the particular cases or do they have external validity? If the theoretical propositions and conceptual perspective are confirmed in one case study (or better: in several case studies), then they can be expected to hold in other cases with similar characteristics. This means that the researcher should try to specify the circumstances under which replication may be expected. The case studies share two characteristics: first, they are located in the past, that is, in the nineteenth and mid-twentieth centuries; and second, the case studies are from the societal domain of transportation. This begs the following two questions: do the conclusions and conceptual perspective also hold for recent times and the future, and do they hold for societal domains other than transportation?

The time dimension could be important, because the network of social groups in the nineteenth century worked differently from that in the late twentieth century. Social groups involved in technological development have become increasingly specialised and differentiated in the course of the twentieth century. During most of the nineteenth century, technical innovation was craft based (for example, the building of wooden sailing ships). In the absence of drawings and scientific calculations, there was little possibility of predicting beforehand whether a new technology would work or serve its intended purpose. The only way to see if a new kind of artefact would work was to build it (McGee, 1999). While technology was not an explicit object of management in the nineteenth century, this changed in the late nineteenth and early twentieth centuries with the emergence of R&D laboratories and state-funded technical programmes. In many sectors, technological development became increasingly science based and government sponsored. To build iron steamships in the late nineteenth century, special schools were set up to train naval architects. Shipbuilding gradually became an applied science. The role of designers and engineers in technological development increased as these social groups professionalised in the second half of the nineteenth century, creating technical societies, professional organisations and technical schools and training institutes. With the emergence of R&D laboratories at the end of the nineteenth century, the position of engineers and researchers was institutionalised inside firms.

Changes also occurred in the position and involvement of public authorities. During much of the nineteenth century the prevailing political culture was liberalism, which meant low involvement of public authorities in public life. There was no explicit technology policy. If subsidies were given, they were driven by other motives (for example, Imperial policy in the case of mail subsidies for steamers). In the late nineteenth and early twentieth centuries, public investments in science (and education) increased, but there was no explicit technology policy. In the context of reform movements and the social question, there was a strong expansion of public administration at all levels (municipal, state, federal). Public authorities took greater responsibility for regulating society, and public policies thus began to influence the selection environment. In Chapter 5 we saw how streets and traffic became more strongly regulated, in effect creating a positive selection environment for the automobile. After the Second World War, public authorities also began to develop policies aimed at technology push. There was stronger involvement in the stimulation of innovation, both through investments in science ('science the endless frontier') and through large technological projects. In the 1945–70 period there was intensified cooperation among government, universities and large companies (the military–industrial complex). In Chapter 6 we saw how governments became involved in many ways in aviation, both through regulation (for example, safety boards, air traffic control) and through innovation (for example, investments in R&D laboratories, procurement programmes). Public involvement in technological innovation remained substantial in the 1980s and 1990s, because it was seen as a crucial driver for economic growth.

Changes also occurred with regard to the role of the public and popular culture. With the increase of public media (newspapers, radio, magazines, television) this role has increased since the late nineteenth century. Many public media reported favourably about new technologies, stimulating the movement of technological utopianism between 1880 and 1930. This movement coincided with the rise of a new popular culture (valuing such things as entertainment, excitement, fun and adventure) and with the Second Industrial Revolution. In Chapter 4 we saw how the steamship transition ended with an innovation race between Britain and Germany to build the largest and fastest superliners. Ships such as the *Mauretania* and *Titanic* excited much public interest. In the transition to automobiles popular culture also played an important role. Car racing enjoyed great public popularity and touring linked up with the cultural values of excitement and adventure. In the emergence of aircraft and aviation, popular culture also played a role. Aircraft enjoyed wide cultural popularity from the 1910s to the 1940s, the so-called 'winged gospel'. Thus, we see the increasing role of public culture and mass media in the three cases. But after 1970 the attitude towards technology changed dramatically, with the rise of a counterculture and societal protest groups which saw technology as the cause of many problems. There were societal concerns about negative side-effects (on the environment, on cities), technocracy and the overwhelming power of technology (for example, Herbert Marcuse's 'one-dimensional man'). In response, new decision-making procedures emerged. Societal groups, citizens and other stakeholders were increasingly involved in the management of system-building projects.

In sum, the network of social groups has changed over time. But do these changes affect the conclusions and conceptual perspective? Yes and no. Yes, because they do affect the mechanisms identified in Section 7.4. The expansion and professionalisation of research and development will have stimulated the mechanism of spatial and geographical proliferation of technologies. The speed of circulation of techno-scientific knowledge has increased substantially in the twentieth century (also through conferences and journals). Researchers and designers follow each other's work and imitate successes. The mechanism of taking large steps in technical innovation (with the risk of failure) may also have become more important. There have been plenty of examples where an R&D-driven innovation push has resulted in too large steps, which resulted in failure. One such example is the failure of the R&D-driven breakthrough trajectory in American wind turbines, and the success of the bottom-up gradual trajectory in Danish wind turbines (Garud and Karnøe, 2003). The increasing involvement of public authorities has made the mechanism of strategic games and innovation races more prominent. In the last two decades countries have often competed for the same strategic technologies (for example, biotechnology, information technology, life sciences). Also, policy support in the creation of niches has increased in the second half of the twentieth century, either through funding of universities or demonstration projects of desirable technologies (for example, sustainable technologies). The increasing role of public culture has increased the prominence of hypes and bandwagon effects. Recent examples are the Internet and third-generation mobile telephony. In addition, the role of wide cultural visions and expectations has increased in the twentieth century. Widely shared visions create legitimation for public investments and stimulation of particular innovations.

While the secular changes in social groups do to some extent influence the mechanisms in system innovations, this influence is far smaller for the other conclusions and conceptual propositions. The general claim that system innovations come about through the alignment of processes at multiple levels is not influenced by the constellation of social groups. Nor are the patterns strongly influenced by the secular changes. The precise trajectory of niche accumulation (which niche first, which niche next) may be influenced by secular changes, but not the general pattern that diffusion takes place through niche accumulation. The same goes for co-evolution of technologies, the fit–stretch pattern in the co-evolution of form and function and the four phases. So I conclude that the multi-level perspective and the patterns also hold for the present.

The reason for this generalisability was mentioned in Chapter 1. The research focus was on long-term processes and patterns, not on the micro level of individual interactions. Local interactions add up to patterns on a more aggregated level. The focus was on the aggregated patterns rather than the local interactions. In other words , I used an analytical approach, which worked 'from the outside in', from a very general multi-level perspective, to patterns and then to mechanisms. Another approach would have been to work 'from the inside out' looking in more detail at actors, how they try to navigate the transition, and how they find their way through searching and learning, involving power struggles, controversies, debates and so on. If I had used such an approach, the findings would have been strongly influenced by the constellation and network of social groups. Then it would have been more difficult to generalise for different time periods.

All three case studies are from the transport domain. Can the findings be generalised to other societal domains? It is difficult to answer this question in a general way, but there are several specificities in the transport domain to which analysts should pay attention if they want to apply my conclusions and findings to other domains.

Transport is characterised by much heterogeneity in user needs and markets. There are a range of larger and smaller transport markets, for example, mail, Army, Navy, freight transport (piece goods, general cargo), personal transportation, different routes and distances. Because there are so many different markets with different selection criteria, it is easy for novelties to find particular niches that suit their characteristics. In domains where there is less heterogeneity, there may be less opportunity for novelties to find a foothold.

In the transport domain the influence of external landscape-level developments is prominent. Wars had a great influence, because ships, cars and planes were used as defence and attack weapons, to transport troops and for logistics. Transport was also influenced by liberalisation and internationalisation, which affected trade patterns. Cultural values such as speed and freedom also influenced transport. Because of this characteristic the case studies showed clearly how diffusion and breakthrough of new technologies depended on external landscape developments, alongside internal drivers. But there may be domains where external developments play a less prominent role and internal drivers are more important. For instance, in domains such as materials supply or natural resources, external cultural influences may be less important. Thus, the importance of external circumstances and internal drivers may differ per case study and domain.

As part of the proposition of phases in system innovations, it was suggested that diffusion and replacement are slow processes. Although this proposition was corroborated by the case studies, it may be related to the specific characteristics of transportation, in particular the presence of many different market niches with slightly different selection criteria, large sunk investments in transport infrastructures and production systems, and capital intensity of transport technologies. When many market niches have to be captured successively, diffusion and replacement take time. Sunk investments in infrastructures provide inertia, and new infrastructures take much time and money to be built. Users will not buy expensive new technologies until investments in old technologies are written off. Likewise, firms do not want to shift to other technologies, until existing production lines are written off. This means that the proposition about slow diffusion and replacement may not always hold. In societal domains, which depend less on infrastructure, with fewer capital-intensive technologies and with more homogeneous markets, diffusion and replacement may proceed faster once the price/performance ratio of the new technology is higher than that of the old one.

The transport domain is characterised by a particular kind of technology, complex assembled products with systemic infrastructure requirements. Artefacts such as ships, aircraft and automobiles consist of many components and subsystems. In such complex artefacts it is likely that the pattern of co-evolution of multiple technologies (for example, components, subsystems, borrowing of technical elements) is more important. Tushman and Rosenkopf (1992) distinguish four kinds of technologies:

1. non-assembled products (for example, aluminium, cement, flat glass, paper, fibres, petroleum, springs and steel);
2. simple assembled products (for example, stoves, hoses, cans, skis, containers, guns, escapements and balance wheels are made up of distinct subsystems that are combined or fit together);
3. complex-assembled products (for example, watches, automobiles and systems are made up of distinct subsystems that interact with each other); and
4. open systems or large technical systems (for example, electricity networks, telephone networks, computer networks and railroads).

The pattern of co-evolution of multiple technologies is likely to be more important for open systems and complex-assembled products than for non-assembled and simple-assembled products. Another consequence is that complex technologies make it possible for system innovations to start with a component change, and then cascade upwards (one of the subroutes of the technological substitution route in Section 7.3). This particular dynamic is less likely for non-assembled and simple-assembled products.

Another characteristic of the transport domain is that there are clear dominant artefacts (for example, cars, planes, ships) around which other aspects of the socio-technical system are arranged (for example, gas stations, roads, traffic lights around cars). This means that technological substitution can easily be observed, as well as the four phases in transition. Furthermore, it is relatively easy to distinguish between stable situations and situations in flux. But there are also societal domains where it is not easy to pinpoint a single dominant technology, for example, agriculture and farming, the medical sector, mining and retailing. In these domains there is no one, distinct artefact that one can follow through several phases and then see it replace the existing artefact. In these domains the dynamic of change may be more gradual and involve many technical changes throughout the system. I have already suggested in Section 7.3 that large technical systems may follow a particular transition route.

Thus, one can specify how aspects of the conclusions may have to be changed for different domains. But most findings seem to hold pretty well. The main reason for this is that the conclusions and patterns are based on generally applicable theoretical considerations of alignment and de-alignment and interlocking of multiple processes. Such a process-oriented approach is necessary for complex topics of analysis, such as socio-technical systems.

Bibliography

Abernathy, William, Kim B. Clark and Alan M. Kantrow (1983), *Industrial Renaissance: Producing a Competitive Future for America*, New York: Basic Books.

Abernathy, W.J. and K.B. Clark (1985), 'Innovation: mapping the winds of creative destruction', *Research Policy*, **14** (1), 3–22.

Abt, Dietmar (1998), *Die Erklärung der Technikgenese des Elektroautomobils*, Frankfurt am Main: Peter Lang.

Akrich, Madeleine (1992), 'The de-scription of technical objects', in Bijker and Law (eds), pp. 205–24.

Akrich, Madeleine and Bruno Latour (1992), 'A summary of a convenient vocabulary for the semiotics of human and nonhuman assemblies', in Bijker and Law (eds), pp. 259–64.

Allen, P.M. (2001), 'A complex systems approach to learning and adaptive networks', *International Journal of Innovation Management*, **5** (2), 149–80.

Anderson, P. (1999), 'Complexity theory and organization science', *Organization Science*, **10** (3), 216–32.

Anderson, P. and M. Tushman (1990), 'Technological discontinuities and dominant designs: a cyclical model of technological change', *Administrative Science Quarterly*, **35**, 604–33.

Arrow, K. (1962), 'The economic implications of learning by doing', *Review of Economic Studies*, **29**, 155–73.

Arthur, Brian (1988), 'Competing technologies: an overview', in Giovanni Dosi, Chris Freeman, Richard Nelson, Gerald Silverberg and Luc Soete (eds), *Technical Change and Economic Theory*, London: Pinter, pp. 590–607.

Ayres, Robert (1989), *Technological Transformations and Long Waves*, Research Report-89-1, Laxenburg: International Institute for Applied Systems Analysis.

Baldwin, C.Y. and K.B. Clark (1997), 'Managing in an age of modularity', *Harvard Business Review*, September–October, 84–93.

Baldwin, Peter C. (1999), *Domesticating the Street: The Reform of Public Space in Hartford, 1850–1930*, Columbus, OH: Ohio State University Press.

Bardou, Jean-Pierre, Jean-Jacques Chanaron, Patrick Fridenson and James

M. Laux (1982), *The Automobile Revolution: The Impact of an Industry*, Chapel Hill, NC: University of North Carolina Press.

Barker, T.C. (1993), 'Slow progress: forty years of motoring research', *Journal of Transport History*, **14**, 142–65.

Barker, Theo C. and Dorian Gerhold (1993), *The Rise and Rise of Road Transport, 1700–1990*, Basingstoke: Macmillan.

Basalla, George (1988), *The Evolution of Technology*, Cambridge: Cambridge University Press.

Baum, Joel A.C. and Jitendra V. Singh (1994), *Evolutionary Dynamics of Organizations*, New York, Oxford: Oxford University Press.

Belasco, Warren J. (1997), *Americans on the Road: From Autocamp to Motel, 1910–1945*, Baltimore, MD and London, UK: Johns Hopkins University Press.

Beniger, James R. (1986), *The Control Revolution: Technological and Economic Origins of the Information Society*, Cambridge, MA and London, UK: Harvard University Press.

Benson, E. (2000), 'Suspicious allies: wartime aviation developments and the Anglo-American international airline rivalry, 1939–1945', *History and Technology*, **17**, 21–42.

Berger, Peter L. and Thomas Luckman (1967), *The Social Construction of Reality*, New York: Doubleday Anchor.

Bertels, K. (1973), *Geschiedenis Tussen Struktuur en Evenement: Een Methodologies en Wijsgerig Onderzoek*, Amsterdam: Wetenschappelijke Uitgeverij BV.

Biggs, Lindy (1996), *The Rational Factory: Architecture, Technology and Work in America's Age of Mass Production*, Baltimore, MD and London, UK: Johns Hopkins University Press.

Bijker, Wiebe E. (1995), *Of Bicycles, Bakelites and Bulbs: Towards a Theory of Sociotechnical Change*, Cambridge, MA and London, UK: MIT Press.

Bijker, Wiebe E. and John Law (eds) (1992), *Shaping Technology/Building Society: Studies in Sociotechnical Change*, Cambridge, MA and London, UK: MIT Press.

Bilstein, Roger E. (1995), 'Air travel and the travelling public: the American experience, 1920–1970', in William F. Trimble (ed.), *From Airships to Airbus: The History of Civil and Commercial Aviation, Volume 2: Pioneers and Operations*, Washington, DC and London, UK: Smithsonian Institution Press, pp. 91–111.

Bonaccorsi, A. and P. Giuri (2000), 'When shakeout doesn't occur: the evolution of the turboprop engine industry', *Research Policy*, **29**, 847–70.

Bourdieu, Pierre (1977), *Outline of a Theory of Practice*, Cambridge: Cambridge University Press.

Bower, J.L. and C.M. Christensen (1995), 'Disruptive technologies: catching the wave', *Harvard Business Review*, January–February, 43–53.

Bowers, B. (1995), 'Electricity', in Ian McNeil (ed.), *An Encyclopedia of the History of Technology*, London and New York: Routledge, pp. 350–87.

Braudel, F. (1958), 'Histoire et sciences sociales: la longue durée', *Annales*, **13**, 725–53.

Braudel, Fernand (1976), *The Mediterranean and the Mediterranean World in the Age of Philip II*, New York: Harper & Row.

Braudel, Fernand (1985a), *The Structures of Everyday Life: The Limits of the Possible; Civilization and Capitalism, 15th–18th Century,* Vol. 1, London: Fontana.

Braudel, Fernand (1985b), *The Wheels of Commerce: Civilization and Capitalism, 15th–18th Century*, Vol. 2, London: Fontana.

Braudel, Fernand (1985c), *The Perspective of the World: Civilization and Capitalism, 15th–18th Century*, Vol. 3, London: Fontana.

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

Broeze, Frank J.A. (1977), 'Rederij', in Frank J.A. Broeze, J.R. Bruijn and F.S. Gaastra, *Maritieme Geschiedenis der Nederlanden: deel 3: Achttiende Eeuw en Eerste Helft Negentiende Eeuw, van ca 1680 tot 1850–1870*, Bussum: De Boer Maritiem, pp. 92–141.

Broeze, F.J.A. (1982), 'The international diffusion of ocean steam navigation: the myth of the retardation of Netherlands steam navigation to the East Indies', *Economisch- en Sociaal-Historisch Jaarboek*, **45**, 77–95.

Bruland, K. and K. Smith (2000), 'Technological transitions in history and theory', paper presented at the international conference 'Transitions to sustainability through system innovation', Twente University, Enschede, 4–6 July, 2002.

Bürgle, Klaus and Richard von Frankenberg (1964), *Autotypen Gestern und Heute*, Stuttgart: Union Verlag.

Burns, Thomas R. and Helena Flam (1987), *The Shaping of Social Organization: Social Rule System Theory with Applications*, London: Sage.

Calhoun, Daniel (1973), *The Intelligence of a People*, Princeton, NJ: Princeton University Press.

Callon, M. (1980), 'The state and technical innovation: a case-study of the electrical vehicle in France', *Research Policy*, **9**, 358–76.

Callon, Michel (1991), 'Techno-economic networks and irreversibility', in John Law (ed.), *A Sociology of Monsters, Essays on Power, Technology and Domination*, London: Routledge, pp. 132–61.

Callon, Michel (ed.) (1998), *The Laws of the Market*, Oxford: Blackwell.

Callon, Michel (1999), 'Actor–network theory: the market test', in John Law and John Hassard (eds), *Actor Network Theory and After*, Oxford: Blackwell, pp. 181–95.

Callon, M., P. Laredo and V. Rabeharisoa (1992), 'The management and evaluation of technological programs and the dynamics of techno-economic networks: the case of the AFME', *Research Policy*, **21** (3), 215–36.

Campbell, Donald T. (1974), 'Evolutionary epistemology', in P.A. Schlipp (ed.), *The Library of Living Philosophers, Vol. XIV, Book I: The Philosophy of Karl Popper*, La Salle, IL: Open Court, pp. 415–63.

Carlsson, B. (ed.) (1997), *Technological Systems and Industrial Dynamics*, Boston, MA, Dordrecht, London: Kluwer Academic Publishers.

Carlsson, B. (2002), *Technological Systems in the Bio Industries: An International Study*, Boston, MA: Kluwer Academic Publishers.

Carlsson, B. and S. Jacobsson (1993), 'Technological systems and economic policy: the diffusion of factory automation in Sweden', *Research Policy*, **23** (3), 235–48.

Carlsson, B. and R. Stankiewicz (1991), 'On the nature, function and composition of technological systems', *Journal of Evolutionary Economics*, **1**, 93–118.

Ceruzzi, Paul (1986), 'An unforeseen revolution: computers and expectations, 1935–1980', in Joseph J. Corn (ed.), *Imagining Tomorrow: History, Technology and the American Future*, Cambridge, MA: MIT Press, pp. 188–201.

Chant, Christopher (1986), *De Geschiedenis van de Luchtvaart*, Alphen aan de Rijn: Atrium.

Chant, Colin (1999), 'The second industrial revolution and the rise of modern urban planning', in Goodman and Chant (eds), pp. 121–77.

Christensen, Clay (1997), *The Innovator's Dilemma: When New Technologies Cause Great Firms to Fail*, Boston, MA: Harvard Business School Press.

Christensen, C., T. Craig and S. Hart (2001), 'The age of technology: the great disruption', *Foreign Affairs*, **80** (2), 80–95.

Clark, K.B. (1985), 'The interaction of design hierarchies and market concepts in technological evolution', *Research Policy*, **14** (5), 235–51.

Constant, Edward W. (1980), *The Origins of the Turbojet Revolution*, Baltimore, MD and London, UK: Johns Hopkins University Press.

Constant, Edward W. II (1987), 'The social locus of technological practice: community, system or organization', in Wiebe E. Bijker, Thomas P. Hughes and Trevor Pinch (eds), *The Social Construction of Technological Systems: New Directions in the Sociology and History of Technology*, Cambridge, MA: MIT Press, pp. 223–42.

Conway, E.M. (2001), 'The politics of blind landing', *Technology and Culture*, **42**, 81–106.

Coombs, Rod, Ken Green, Albert Richards and Vivian Walsh (eds) (2001), *Technology and the Market: Demand, Users and Innovation*, Cheltenham, UK and Northampton, MA, USA: Edward Elgar.

Coombs, Rod, Paolo Saviotti and Vivian Walsh (1992), 'Technology and the firm: the convergence of economic and sociological approaches', in Rod Coombs, Paolo Saviotti and Vivian Walsh (eds), *Technological Change and Company Strategies: Economic and Sociological Perspectives*, London: Academic Press, pp. 1–24.

Corn, Joseph J. (1983), *The Winged Gospel: America's Romance with Aviation, 1900–1950*, New York: Oxford University Press.

Correljé, Aad and Geert Verbong (2004), *The Transition from Coal to Gas: Radical Change of the Dutch Gas System*, in Boelie Elzen, Frank W. Geels and Ken Green (eds), *System Innovation and the Transition to Sustainability*, Cheltenham, UK and Northampton, MA, USA: Edward Elgar.

Coutard, Olivier (ed.) (1999), *The Governance of Large Technical Systems*, London: Routledge.

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

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

Craig, Robin (1980), *The Ship: Steam Tramps and Cargo Liners, 1850–1950*, London: Her Majesty's Stationery Office.

Cronberg, Tarja and Knut H. Sørensen (1995), 'Similar concerns, different styles? A note on European approaches to the social shaping of technology', in Tarja Cronberg and Knut H. Sørensen (eds), *Similar Concerns, Different Styles? Technology Studies in Western Europe*, Luxembourg: Office for Official Publications of the European Communities, pp. 1–23.

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

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

Davis, D.F. (1995), 'North American urban mass transit, 1890–1950: what if we thought about it as a type of technology?', *History and Technology*, **12**, 309–26.

Dawson, Virginia P. (1995), 'The American turbojet industry and British competition: the mediating role of government research', in William M. Leary (ed.), *From Airships to Airbus: The History of Civil and Commercial Aviation, Volume 1: Infrastructure and Environment*, Washington, DC and London, UK: Smithsonian Institution Press, pp. 27–150.

Dierikx, Marc and Bram Bouwens (1997), *Building Castles of the Air: Schiphol Amsterdam and the Development of Airport Infrastructure in Europe, 1916–1996*, The Hague: Sdu Publishers.

Dierikx, M. and J. Schot (1998), 'Technische Ontwikkeling van Schiphol, 1916–1970', Report, Eindhoven University of Technology, Eindhoven.

DiMaggio, Powell (1990), 'Cultural aspects of economic action and organization', in Roger Friedland and A.F. Robertson (eds), *Beyond the Marketplace*, New York: Aldine de Gruyter, pp. 113–36.

DiMaggio, Powell and Sharon Zukin (1990), 'Introduction', in Sharon Zukin and Powell DiMaggio (eds), *Structures of Capital*, Cambridge: Cambridge University Press, pp. 1–36.

Dirkzwager, J.M. (1978), 'Schepen', in R. Baetens, Ph.M. Bosscher and H. Rechlin (eds), *Maritieme Geschiedenis der Nederlanden: deel 4: Tweede Helft Negentiende Eeuw en Twintigste Eeuw, van 1850–1870 tot ca 1970*, Bussum: De Boer Maritiem, pp. 13–52.

Dirkzwager, J.M. (1993), 'Scheepsbouw', in Harry W. Lintsen (ed.), *Geschiedenis van de Techniek in Nederland: De Wording van een Moderne Samenleving, 1800–1890*, Vol. IV, Zutphen: Walburg Pers, pp. 66–102.

Disco, C., A. Rip and B. Van der Meulen (1992), 'Technical innovation and the universities: divisions of labour in cosmopolitan regimes', *Social Science Information*, **31**, 465–507.

Doloreux, D. (2002), 'What we should know about regional systems of innovation', *Technology in Society*, **24**, 243–63.

Dosi, G. (1982), 'Technological paradigms and technological trajectories: a suggested interpretation of the determinants and directions of technical change', *Research Policy*, **6**, 147–62.

Douglas, Deborah G. (1995), 'Airports and systems of airports: airports and urban development before World War II', in William M. Leary (ed.), *From Airships to Airbus: The History of Civil and Commercial Aviation, Volume 1: Infrastructure and Environment*, Washington, DC and London, UK: Smithsonian Institution Press, pp. 55–84.

Du Gay, Paul, Stuart Hall, Linda Janes, Hugh MacKay and Keith Negus (1997), *Doing Cultural Studies: The Story of the Sony Walkman*, London: Sage.

Durkheim, Émile (1949), *The Division of Labour in Society*, Glencoe, IL: Free Press.

Eldridge, C. (2000), 'Electronic eyes for the allies: Anglo-American cooperation on radar development during World War II', *History and Technology*, **17**, 1–20.

Elias, Norbert (1982), *The Civilizing Process: State Formation and Civilization*, Oxford: Basil Blackwell (original publication in German, *Ueber den Prozess der Zivilisation: Soziogenetische und Psychogenetische Untersuchungen*, 1939).

Emery, Fred (1993), 'Characteristics of socio-technical systems', in Trist and Murray (eds), pp. 157–86.

Engel, J.A. (2000), ' "We are not concerned who the buyer is": engine sales and Anglo-American security at the dawn of the jet age', *History and Technology*, **17**, 43–67.

Etzkowitz, H. and L. Leydesdorff (2000), 'The dynamics of innovation: from national systems and "Mode 2" to a triple helix of university–industry–government relations', *Research Policy*, **29** (2), 109–23.

Faragher, John M., Mari J. Buhle, Daniel Czitrom and Susan H. Armitage (1997), *Out of Many: A History of the American People*, 2nd edn, Englewood Cliffs, NJ: Prentice-Hall.

Fischer, Claude S. (1992), *America Calling: A Social History of the Telephone to 1940*, Berkeley, CA: University of California Press.

Fleck, J. (1993), 'Configurations: crystallizing contingency', *International Journal of Human Factors in Manufacturing*, **3**, 15–36.

Fleck, Jamie (2000), 'Artefact ↔ activity: the coevolution of artefacts, knowledge and organization in technological innovation', in John Ziman (ed.), *Technological Innovation as an Evolutionary Process*, Cambridge: Cambridge University Press, pp. 248–66.

Fletcher, R.A. (1910), *Steam-ships: The Story of their Development to the Present Day*, London: Sidgwick & Jackson.

Flink, James J. (1990), *The Automobile Age*, Cambridge, MA: MIT Press.

Fogelberg, Hans (1997), *The Electric Car Controversy: A Social-Constructivist Interpretation of the California Zero-Emission Vehicle Mandate*, Göteborg: Chalmers University of Technology Press.

Foster, Richard N. (1986), *Innovation: The Attacker's Advantage*, London: Macmillan.

Freeman, Chris (1988), 'Japan: a new national system of innovation', in Giovanni Dosi, Chris Freeman, Richard Nelson, Gerald Silverberg and Luc Soete (eds), *Technical Change and Economic Theory*, London: Frances Pinter, pp. 330–48.

Freeman, Chris and Francisco Louçã (2001), *As Time Goes By: From the Industrial Revolutions to the Information Revolution*, Oxford: Oxford University Press.

Freeman, Chris and Carlotta Perez (1988), 'Structural crisis of adjustment, business cycles and investment behaviour', in Giovanni Dosi, Chris Freeman, Richard Nelson, Gerald Silverberg and Luc Soete (eds),

Technical Change and Economic Theory, London: Frances Pinter, pp. 38–66.
Freeman, Chris and Luc Soete (1997), *The Economics of Industrial Innovation*, 3rd edn, Cambridge, MA: MIT Press.
Frenken, K., P.P. Saviotti and M. Trommetter (1999), 'Variety and niche creation in aircraft, helicopters, motorcycles and minicomputers', *Research Policy*, **28**, 469–88.
Gardiner, J.P. (1984), 'Design trajectories for airplanes and automobiles', in Chris Freeman (ed.), *Design, Innovation and Long Cycles in Economic Development*, London: Frances Pinter, pp. 121–42.
Garrat, G.R.M., David C. Goodman and Colin A. Russel (1973), *Science and Engineering*, Unit 14 and 15, Bletchley: Open University Press.
Garud, R. and P. Karnøe (2003), 'Bricolage versus breakthrough: distributed and embedded agency in technological entrepreneurship', *Research Policy*, **32**, 277–300.
Garud, R. and M.A. Rappa (1994), 'A socio-cognitive model of technology evolution: the case of cochlear implants', *Organization Science*, **5** (3), 344–62.
Geels, F.W. (2002), 'Technological transitions as evolutionary reconfiguration processes: a multi-level perspective and a case-study', *Research Policy*, **31** (8/9), 1257–74.
Geels, F.W. (2004a), 'From sectoral systems of innovation to socio-technical systems: insights about dynamics and change from sociology and institutional theory', *Research Policy* (forthcoming).
Geels, Frank W. (2004b), 'Understanding system innovations: a critical literature review and a conceptual synthesis', in Boelie Elzen, Frank W. Geels and Ken Green (eds), *System Innovation and the Transition to Sustainability: Theory, Evidence and Policy*, Cheltenham, UK and Northampton, MA, USA: Edward Elgar (forthcoming).
Geels, Frank W. and Rene Kemp (2000), *Transities vanuit sociotechnisch perspectief*, Report for the Dutch Ministry of Environment, Enschede: Universiteit Twente, and Maastricht: MERIT.
Georgano, Nick (1985), *Bilen 1886–1930*, Gothenborg: Nordbok.
Geyer, A. and Davies, A. (2000), 'Managing project-system interfaces: case-studies of railway projects in restructured UK and German markets', *Research Policy*, **29** (7–8), 991–1013.
Giddens, Anthony (1984), *The Constitution of Society: Outline of the Theory of Structuration*, Berkeley, CA: University of California Press.
Gilbert, Glen A. (1973), *Air Traffic Control: The Uncrowded Sky*, Washington, DC: Smithsonian Institution Press.
Gilfillan, S.C. (1935), *Inventing the Ship: A Study of the Inventions Made in Her History between Floating Log and Rotorship*, Chicago, IL: Follett.

Goffmann, Erving (1961), *Encounters: Two studies in the sociology of Interaction*, Indianapolis: Bobbs-Merrill.
Goodman, David (1999), 'Cities of the industrial revolution (to 1870)', in Goodman and Chant (eds), pp. 1–30.
Goodman, David and Colin Chant (1999), *European Cities and Technology: Industrial to Post-Industrial City*, London: The Open University, Routledge.
Graham, C. (1956), 'The ascendency of the sailing ship, 1850–1885', *Economic History Review*, **9**, 74–88.
Granovetter, M. (1985), 'Economic action and social structure: the problem of embeddedness', *American Journal of Sociology*, **91**, 481–510.
Green, Ken (1992), 'Creating demand for biotechnology: shaping technologies and markets', in Rod Coombs, Paolo Saviotti and Vivian Walsh (eds), *Technological Change and Company Strategies: Economic and Sociological Perspectives*, London: Academic Press, pp. 164–84.
Griffith, T.L. and D.J. Dougherty (2001), 'Beyond socio-technical systems: introduction to the special issue', *Journal of Engineering and Technology Management*, **18** (3–4), 207–18.
Grübler, A. (1991), 'Diffusion: long-term patterns and discontinuities', *Technological Forecasting and Social Change*, **39** (1–2), 159–80.
Grübler, Arnold (1998), *Technology and Global Change*, Cambridge: Cambridge University Press.
Grübler, Arnold and Nebosja Nakićenović (1991), *Long Waves, Technology Diffusion, and Substitution*, Research Report-91-17, Laxenburg: International Institute for Applied Systems Analysis.
Gunston, Bill (1997), *The Development of Jet and Turbine Aero Engines*, 2nd edn, Sparkford: Patrick Stephens.
Hall, Peter (1998), *Cities in Civilization: Culture, Innovation, and Urban Order*, London: Phoenix Giant.
Hård, Michael and Andreas Knie (1994), 'The ruler of the game: the defining power of the standard automobile', in Knut H. Sørensen (ed.), *The Car and Its Environments: The Past, Present and Future of the Motor Car in Europe*, Brussels: European Commission, pp. 137–58.
Harley, C. Knick (1971), 'The shift from sailing ships to steamships, 1850–1890: a study in technological change and its diffusion', in Donald McCloskey (ed.), *Essays on a Mature Economy: Britain after 1840*, London: Methuen, pp. 215–34.
Harley, C. Knick (1973), 'On the persistence of old techniques: the case of North American wooden shipbuilding', *Journal of Economic History*, **33** (2), 372–98.
Harley, C. Knick (1985), 'Aspects of the economics of shipping', in Lewis R. Fischer and Gerals E. Panting (eds), *Change and Adaptation in*

Maritime History: The North Atlantic Fleets in the Nineteenth Century, St Johns, NL: Maritime History Group, Memorial University of Newfoundland, pp. 169–86.

Harley, C.K. (1988), 'Ocean freight rates and productivity 1740–1913: the primacy of mechanical invention reaffirmed', *Journal of Economic History*, **48**, 851–76.

Harrison, Richard T. (1990), *Industrial Organisation and Changing Technology in UK Shipbuilding: Historical Developments and Future Implications*, Aldershot: Avebury and Gower.

Henderson, R.M. and K.B. Clark (1990), 'Architectural innovation: the reconfiguration of existing product technologies and the failure of established firms', *Administrative Science Quarterly*, **35**, 9–30.

Heppenheimer, T.A. (1995), *Turbulent Skies: The History of Commercial Aviation*, New York: John Wiley.

Hilton, G.W. (1969), 'Transport technology and the urban pattern', *Journal of Contemporary History*, **4**, 123–35.

Hobday, M. (1998), 'Product complexity, innovation and industrial organisation', *Research Policy*, **26**, 689–710.

Hobday, M., H. Rush and J. Tidd (2000), 'Innovation in complex products and systems', *Research Policy*, **29** (7–8), 793–804.

Hodgson, G.M. (1998), 'The approach of institutional economics', *Journal of Economic Literature*, **36**, 166–92.

Holling, C.S. (2001), 'Understanding the complexity of economic, ecological and social systems', *Ecosystems*, **4**, 390–405.

Hoogma, Remco (2000), 'Exploiting technological niches: strategies for experimental introduction of electric vehicles', PhD thesis, Twente University, Enschede.

Hoogma, Remco, René Kemp, Johan Schot and Bernard Truffer (2002), *Experimenting for Sustainable Transport: The Approach of Strategic Niche Management*, London and New York: Spon.

Hopps, R.H. (1978), 'Aircraft manufacturing and utilization', *Proceedings of the International Conference of Tourism and Air Transport*, Madrid: World Trade Organisation and Geneva: International Air Transport Association, pp. 87–131.

Hounshell, David A. (1984), *From the American System to Mass Production, 1800–1932: The Development of Manufacturing Technology in the United States*, Baltimore, MD and London, UK: Johns Hopkins University Press.

Howells, J. (2002), 'The response of old technology incumbents to technological competition: does the sailing ship effect exist?', *Journal of Management Studies*, **39** (7), 887–906.

Hughes, Thomas P. (1983), *Networks of Power: Electrification in Western Society, 1880–1930*, Baltimore, MD: Johns Hopkins University Press.

Hughes, T.P. (1986), 'The seamless web: technology, science, etcetera, etcetera', *Social Studies of Science*, **16**, 281–92.

Hughes, Thomas P. (1987), 'The evolution of large technological systems', in Wiebe E. Bijker, Thomas P. Hughes and Trevor Pinch (eds), *The Social Construction of Technological Systems: New Directions in the Sociology and History of Technology*, Cambridge, MA: MIT Press, pp. 51–82.

Hughes, Thomas P. (1989), *American Genesis: A Century of Invention and Technological Enthusiasm, 1870–1970*, New York: Viking Press.

Hughes, Thomas P. (1998), *Rescuing Prometheus*, New York: Pantheon.

Hughes, Thomas P. (1994), 'Technological momentum', in Merrit R. Smith and Leo Marx (eds), *Does Technology Drive History? The Dilemma of Technological Determinism*, Cambridge, MA: MIT Press, pp. 101–13.

Hulton Deutsch Collection (1995), *De Beste Foto's uit de 19^{e} eeuw: De Maatschappij van 1850–1918 Gevangen in Honderden Foto's*, Lisse: Rebo Productions.

Hunter, Louis C. (1949), *Steamboats on the Western Rivers: An Economic and Technological History*, Cambridge, MA: Harvard University Press.

Hunter, Louis C. and Lynwood Bryant (1991), *The History of Industrial Power in the United States, 1780–1930*, Vol. 3, *The Transmission Power*, Cambridge, MA: MIT Press.

Islas J. (1997), 'Getting round the lock-in in electricity generating systems: the example of the gas turbine', *Research Policy*, **26**, 49–66.

Jablonka, Eva and John Ziman (2000), 'Biological evolution: processes and phenomena', in John Ziman (ed.), *Technological Innovation as an Evolutionary Process*, Cambridge: Cambridge University Press, pp. 13–26.

Jackson, Gordon (1998), 'Ports, ships and government in the nineteenth and twentieth centuries', in Paul C. Van Royen, Lewis R. Fischer and David M. Williams (eds), *Frutta di Mare: Evolution and Revolution in the Maritime World in the 19th and 20th Centuries*, Amsterdam: Batavian Lion International, pp. 161–76.

Jacobsson, Staffan and Anna Johnson (2000), 'The diffusion of renewable energy technolgy: an analytical framework and key issues for research', *Energy Policy*, **28**, 625–40.

Jamison, A. (1970), *The Steam-Powered Automobile: An Answer to Air Pollution*, Bloomington, IN and London, UK: Indiana University Press.

Jarvis, Adrian (1998), 'The nineteenth-century roots of globalization: some technological considerations', in David J. Starkey and Gelina Harlaftis (eds), *Global Markets: The Internationalization of the Sea Transport*

Industries Since 1850, St Johns, NL: International Maritime Economic History Association, pp. 217–37.
Joerges, Bernward (1988), ‘Large technical systems: concepts and issues’, in Mayntz and Hughes (eds), pp. 9–36.
Josephy Jr., Alvin M. (1963), *Het grootse avontuur van de luchtvaart: Van luchtballon tot straalvliegtuig*, Amsterdam and Brussels: Elsevier.
Kaukiainen, Y. (1992), ‘Coal and canvas: aspects of the competition between steam and sail, 1870–1914’, *International Journal of Maritime History*, **IV** (2), 175–91.
Kaukiainen, Yrjö (1998), ‘International freight markets in the 1830s and 1840s: the experience of a major Finnish shipowner’, in David J. Starkey and Gelina Harlaftis (eds), *Global Markets: The Internationalization of the Sea Transport Industries since 1850*, St Johns, NL: International Maritime Economic History Association, pp. 1–27.
Kemp, R. (1994), ‘Technology and the transition to environmental sustainability. The problem of technological regime shifts’, *Futures*, **26**, 1023–46.
Kemp, René, Arie Rip and Johan Schot (2001), ‘Constructing transition paths through the management of niches’, in Raghu Garud and Peter Karnoe (eds), *Path Dependence and Creation*, Mahwah, NJ: Lawrence Erlbaum, pp. 269–99.
Kemp, R., J. Schot and R. Hoogma (1998), ‘Regime shifts to sustainability through processes of niche formation: the approach of strategic niche management’, *Technology Analysis and Strategic Management*, **10**, 175–96.
Kirsch, David A. (1996), ‘The electric car and the burden of history: studies in, automotive systems rivalry in America, 1890–1996’, PhD thesis, Department of History, Stanford University, Stanford, CA.
Klepper, S. (2002), ‘The capabilities of new firms and the evolution of the US automobile industry’, *Industrial and Corporate Change*, **11** (4), 645–66.
Kline, R. and T. Pinch (1996), ‘Users as agents of technological change: the social construction of the automobile in the rural United States’, *Technology and Culture*, **37**, 763–95.
Kline, Stephen J. and Nathan Rosenberg (1986), ‘An overview of Innovation’, in Ralph Landau and Nathan Rosenberg (eds), *The Positive Sum Strategy: Harnessing Technology for Economic Growth*, Washington, DC: National Academy Press, pp. 275–305.
La Porte, T.R. (1988), ‘The United States air traffic system: increasing reliability in the midst of rapid growth’, in Mayntz and Hughes (eds), pp. 215–44.
La Porte, Todd R. (ed.) (1991), *Social Responses to Large Technical Systems: Control or Anticipation*, Dordrecht: Kluwer Academic Publishers.

Lambert, A.D. (1999), 'Responding to the nineteenth century: the Royal Navy and the introduction of the screw propeller', *History of Technology*, **21**, 1–28.

Landes, David S. (1995), *The Unbound Prometheus: Technological Change and Industrial Development in Western Europe from 1750 to the Present*, Cambridge: Cambridge University Press.

Langlois, R.N. (2002), 'Modularity in technology and organization', *Journal of Economic Behavior and Organization*, **49**, 19–37.

Latour, Bruno (1987), *Science in Action*, Cambridge, MA: Harvard University Press.

Latour, Bruno (1991), 'Society is technology made durable', in John Law (ed.), *A Sociology of Monsters: Essays on Power, Technology and Domination*, London: Routledge, pp. 103–31.

Latour, Bruno (1992), 'Where are the missing masses? The sociology of a few mundane artefacts', in Bijker and Law (eds), pp. 205–24.

Latour, Bruno (1993), *La Clef de Berlin et autres leçons d'un amateur de sciences*, Paris: Éditions la Découverte.

Law, John (1987), 'Technology and heterogeneous engineering: the case of Portuguese expansion', in Wiebe E. Bijker, Thomas P. Hughes and Trevor Pinch (eds), *The Social Construction of Technological Systems: New Directions in the Sociology and History of Technology*, Cambridge, MA: MIT Press, pp. 111–34.

Law, John and Michel Callon (1992), 'The life and death of an aircraft: a network analysis of technical change', in Bijker and Law (eds), pp. 21–52.

Layton, E. (1971), 'Mirror-image twins: the communities of science and technology in 19th century America, *Technology and Culture*, **12** (4), 562–80.

Layton, E. (1976), 'American ideologies of science and engineering', *Technology and Culture*, **17** (4), 688–701.

Leary, William M. (1995), 'Safety in the air: the impact of instrument flying and radio navigation on U.S. commercial air operations between the wars', in William M. Leary (ed.), *From Airships to Airbus: The History of Civil and Commercial Aviation, Volume 1: Infrastructure and Environment*, Washington, DC and London, UK: Smithsonian Institution Press, pp. 97–113.

Leonard-Barton, D. (1988), 'Implementation as mutual adaptation of technology and organization', *Research Policy*, **17** (5), 251–67.

Leonard-Barton, Dorothy (1995), *Wellsprings of Knowledge: Building and Sustaining the Sources of Innovation*, Boston, MA: Harvard Business School Press.

Levinthal, D.A. (1998), 'The slow pace of rapid technological change: gradualism and punctuation in technological change', *Industrial and Corporate Change*, **7** (2), 217–47.

Lewis, Tom (1999), *Divided Highways: Building the Interstate Highways, Transforming American Life*, Harmondsworth, UK: Penguin Books.

Leydesdorff, L. (2000), 'The triple helix: an evolutionary model of innovations', *Research Policy*, **29**, 243–55.

Leydesdorff, L. and H. Etzkowitz (1998), 'The triple helix as a model for innovation studies', *Science and Public Policy*, **25** (3), 195–203.

Lie, Merete and Knut H. Sørensen (eds) (1996), *Making Technology Our Own: Domesticating Technology into Everyday Life*, Oslo: Scandinavian University Press.

Ling, Peter J. (1990), *America and the Automobile: Technology, Reform and Social Change, 1893–1923*, Manchester and New York: Manchester University Press.

Lissoni, Francesco and Stan Metcalfe (1994), 'Diffusion of innovation ancient and modern: a review of the main themes', in Marc Dodgson and Roy Rothwell (eds), *The Handbook of Industrial Innovation*, Aldershot, UK and Brookfield, US: Edward Elgar, pp. 106–41.

Lonati, R. (1972), 'Air transport: catalyst of world tourism', *Air Transport: Catalyst of World Tourism; Proceedings of 28th annual general meeting*, Montreal: International Air Transport Association, pp. 11–16.

Lundvall, Bengt A. (1988), 'Innovation as an interactive process: from user–producer interaction to the national system of innovation', in Giovanni Dosi, Chris Freeman, Richard Nelson, Gerald Silverberg and Luc Soete (eds), *Technical Change and Economic Theory*, London: Pinter, pp. 349–69.

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

Lynn, G.S., J.G. Morone and A.S. Paulson (1996), 'Marketing and discontinuous innovation: the probe and learn process', *California Management Review*, **38**, 8–37.

MacKenzie, Donald (1992), 'Economic and sociological explanations of technical change', in Rod Coombs, Paolo Saviotti and Vivian Walsh (eds), *Technological Change and Company Strategies: Economic and Sociological Perspectives*, London: Academic Press, pp. 25–48.

Maddocks, Melvin (1982), *De Grote Oversteek*, London, Amsterdam: Time-Life.

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

Martin, Paul (2001), 'Great expectations: the construction of markets, products, and user needs during the early development of gene therapy in the USA', in Coombs, Green, Richards and Walsh (eds), pp. 38–67.

Marvin, Carolyn (1988), *When Old Technologies Were New: Thinking About Electric Communication in the Late Nineteenth Century*, Oxford: Oxford University Press.

Mayntz, Renate and Thomas P. Hughes (eds) (1988), *The Development of Large Technical Systems*, Frankfurt: Campus Verlag; and Boulder, CO: Westview.

McGee, D. (1999), 'From craftmanship to draftmanship: naval architecture and the three traditions of early modern design', *Technology and Culture*, **40**, 209–36.

McNeil, Ian (1995), 'Roads, bridges and vehicles', in Ian McNeil (ed.), *An Encyclopedia of the History of Technology*, London and New York: Routledge, pp. 431–73.

McShane, Clay (1994), *Down the Asphalt Path: The Automobile and the American City*, New York: Columbia University Press.

Miller, R., M. Hobday, T. Leroux-Demers and X. Olleros (1995), 'Innovation in complex system industries: the case of flight simulation', *Industrial and Corporate Change*, **4** (2), 363–400.

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

Mokyr, Joel (1990), *The Lever of Riches*, New York: Oxford University Press.

Mokyr, Joel (2000), 'Evolutionary phenomena in technological change', in John Ziman (ed.), *Technological Innovation as an Evolutionary Process*, Cambridge: Cambridge University Press, pp. 52–65.

Molina, A.H. (1999), 'Understanding the role of the technical in the build-up of sociotechnical constituencies', *Technovation*, **19**, 1–19.

Mom, Gijs (1997), 'Geschiedenis van de auto van morgen: Cultuur en techniek van de elektrische auto', PhD thesis, University of Nijmegen, Deventer: Kluwer Bedrijfsinformatie BV.

Mom, G.P.A. and D.A. Kirsch (2001), 'Technologies in tension: horses, electric trucks, and the motorization of American cities, 1900–1925', *Technology and Culture*, **42**, 489–518.

Mom, Gijs and Vincent van der Vinne (1995), *De elektro-auto: Een paard van Troje?*, Deventer: Kluwer Bedrijfsinformatie BV.

Mom, Gijs, Charley Werff and Ariejan Bos (1997), *De auto: Van avonturen-machine naar gebruiksvoorwerp*, Deventer: Kluwer Bedrijfsinformatie BV.

Moore, William L. (1994), 'Radical innovations: what can be learned from the past?', in Marjorie Adams and Joseph LaCugna (eds), *And Now for Something Completely Different: 'Really' New Products*, Cambridge, MA: Marketing Science Institute, pp. 3–10

Morel, B. and R. Ramanujam (1999), 'Through the looking glass of complexity: the dynamics of organizations as adaptive and evolving systems', *Organization Science*, **10** (3), 278–93.

Mowery, David C. and Nathan Rosenberg (1982), 'Technical change in the commercial aircraft industry, 1925–1975', in Rosenberg (ed.), pp. 163–77.

Mowery, David C. and Nathan Rosenberg (1998), *Paths of Innovation: Technological Change in 20th-century America*, Cambridge: Cambridge University Press.

Nahum, A. (1997), 'Two-stroke or turbine? The aeronautical research committee and British aero-engine development in World War II', *Technology and Culture*, **38**, 312–54.

Nakićenović, N. (1986), 'The automobile road to technological change: diffusion of the automobile as a process of technological substitution', *Technological Forecasting and Social Change*, **29** (4), 309–40.

Nakićenović, N. (1991), 'Diffusion of pervasive systems: a case of transport infrastructures', *Technological Forecasting and Social Change*, **39**, 181–200.

National Research Council (1999), *Our Common Journey. A Transition Toward Sustainability*, Washington, DC: National Academy Press.

Nelson, Richard R. (ed.) (1993), *National Systems of Innovation: A Comparative Analysis*, Oxford: Oxford University Press.

Nelson, R.R. (1994a), 'The co-evolution of technology, industrial structure, and supporting institutions', *Industrial and Corporate Change*, **3** (1), 47–63.

Nelson, Richard R. (1994b), 'Economic growth via the coevolution of technology and institutions', in Loet Leydesdorff and Peter van den Besselaar (eds), *Evolutionary Economics and Chaos Theory: New Directions in Technology Studies*, London: Pinter, pp. 21–32.

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

Nelson, Richard R. (2000), 'Selection criteria and selection processes in cultural evolution theories', in John Ziman (ed.), *Technological Innovation as an Evolutionary Process*, Cambridge: Cambridge University Press, pp. 66–74.

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

Nijholt, Anton and Jan van den Ende (1994), *Geschiedenis van de Rekenkunst: Van Kerfstok tot Computer*, Schoonhoven: Academic Service.

North, Douglas C. (1990), *Institutions, Institutional Change and Economic Performance*, Cambridge: Cambridge University Press.

Nye, David (1990), *Electrifying America: Social Meanings of a New Technology*, Cambridge, MA and London, UK: MIT Press.

Nye, David E. (1998), *Consuming Power: A Social History of American Energies*, Cambridge, MA and London, UK: MIT Press.

O'Rourke, Kevin H. and Jeffrey G. Williamson (1999), *Globalization and History: The Evolution of a Nineteenth-Century Atlantic Economy*, Cambridge, MA and London, UK: MIT Press.

Olleros, F. (1986), 'Emerging industries and the burnout of pioneers', *Journal of Product Innovation Management*, **1**, 5–18.

Olson, David (1994), 'Challenges in "creating customers" via really new products', in Marjorie Adams and Joseph LaCugna (eds), *And Now for Something Completely Different: 'Really' New Products*, Cambridge, MA: Marketing Science Institute, pp. 32–6.

OTA (Office of Technology Assessment) (1982), *Airport and Air Traffic Control System*, Washington, DC: US Government Printing Office .

OTA (Office of Technology Assessment) (1994), *Federal Research and Technology for Aviation*, Washington, DC: US Government Printing Office.

Oudshoorn, Nelly and Trevor Pinch (eds) (2003), *How Users Matter: The Co-Construction of Users and Technology*, Cambridge, MA: MIT Press.

Page, Stephen J. (1999), *Transport and Tourism*, New York: Addison Wesley Longman.

Palmer, R.R. and J. Colton (1992), *A History of the Modern World, since 1815*, 7th edn, New York: McGraw-Hill.

Palmer, Sarah (1985), 'The British shipping industry, 1850–1914', in Lewis R. Fischer and Gerald E. Panting (eds), *Change and Adaptation in Maritime History: The North Atlantic Fleets in the Nineteenth Century*, St Johns, NL: Maritime History Group, Memorial University of Newfoundland, pp. 89–113.

Parsons, T. (1937), *The Structure of Social Action*, New York: McGraw-Hill.

Pattee, Howard H. (ed.) (1973a), *Hierarchy Theory: The Challenge of Complex Systems*, New York: George Braziller.

Pattee, Howard H. (1973b), 'Unsolved problems and potential applications of hierarchy theory', in Pattee (ed.), pp. 129–56.

Pavitt, K. (1984), 'Sectoral patterns of technical change: towards a taxonomy and a theory', *Research Policy*, **13**, 343–73.

Pearsall, A.W.H. (1996), 'Ports and shipping', in Ian McNeil (ed.), *An Encyclopedia of the History of Technology*, London and New York: Routledge, pp. 519–54.

Pinch, Trevor J. and Wiebe E. Bijker (1987), 'The social construction of facts and artifacts: or how the sociology of science and the sociology of technology might benefit each other', in Wiebe E. Bijker, Thomas P. Hughes and Trevor J. Pinch (eds), *The Social Construction of Technological Systems: New Directions in the Sociology and History of Technology*, Cambridge, MA: MIT Press, pp. 17–50.

Pistorius, C.W.I. and J.M. Utterback (1997), 'Multi-mode interaction among technologies', *Research Policy*, **26** (1), 67–84.

Pollard, Sidney and Paul Robertson (1979), *The British Shipbuilding Industry, 1870–1914*, Cambridge, MA and London, UK: Harvard University Press.

Potts, Jason (2000), *The New Evolutionary Microeconomics: Complexity, Competence and Adaptive Behaviour*, Cheltenham, UK and Northampton, MA, USA: Edward Elgar.

Pursell, Caroll (1995), *The Machine in America: A Social History of Technology*, Baltimore, MD and London, UK: Johns Hopkins University Press.

Rae, John B. (1968a), *Climb to Greatness: The American Aircraft Industry, 1920–1960*, Cambridge, MA: MIT Press.

Rae, J.B. (1968b), 'Transportation technology and the problem of the city', *Traffic Quarterly*, **22**, 299–314.

Rammert, W. (1997), 'New rules of sociological method: rethinking technology studies', *British Journal of Sociology*, **48** (2), 171–91.

Ransom, P.J.G. (1995), 'Rail', in Ian McNeil (ed.), *An Encyclopedia of the History of Technology*, London and New York: Routledge, pp. 555–608.

Rao, Hayagreeva and Jitendra V. Singh (2001), 'The construction of new paths: institution-building in the early automobile and biotechnology industries', in Raghu Garud and Peter Karnøe (eds), *Path Dependence and Creation*, Mahwah, NJ: Lawrence Erlbaum, pp. 243–67.

Raskin, Paul, Tariq Banuri, Gilberto Gallopin, Pablo Butman, Al Hammond, Robert Kates and Rob Swart (2002), *Great Transition: The Promise and Lure of the Times Ahead*, Boston, MA: Stockholm Environment Institute.

Rip, A. (1992), 'Between innovation and evaluation: sociology of technology applied to technology policy and technology assessment', *Rivista di Studi Epistemologici e Sociali Sulla Scienza e la Tecnologia* (*RISESST)*, **2**, 39–68.

Rip, Arie (2000), 'There's no turn like the empirical turn', in Peter Kroes, Anthonie Meijers and Carl Mitcham (eds), *The Empirical Turn in the Philosophy of Technology*, Amsterdam: Elsevier Science, pp. 3–17.

Rip, Arie and Aard J. Groen (2001), 'Many visible hands', in Coombs, Green, Richards and Walsh (eds), pp. 12–37.

Rip, Arie and René Kemp (1998), 'Technological change', in Steven Rayner and Elly L. Malone (eds), *Human Choice and Climate Change*, Vol. 2, Columbus, OH: Battelle Press, pp. 327–99.

Rip, Arie and Johan W. Schot (2002), 'Identifying loci for influencing the dynamics of technological development', in Knut H. Sørensen and Robin Williams (eds), *Shaping Technology, Guiding Policy: Concepts, Spaces and Tools*, Cheltenham, UK and Northampton, MA, USA: Edward Elgar, pp. 155–72.

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

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

Rosenberg, N. (1986), 'The impact of technological innovation: a historical view', in R. Landau and N. Rosenberg (eds), *The Positive Sum Strategy: Harnessing Technology for Economic Growth*, Washington, DC: National Academy Press, pp. 17–32.

Rosenkopf, Lori and Michael L. Tushman (1994), 'The coevolution of technology and organization', in Baum and Singh (eds), pp. 403–24.

Rotmans, J., R. Kemp and M. van Asselt (2001), 'More evolution than revolution: transition management in public policy', *Foresight*, **3** (1), 15–31.

Rotmans, Jan, René Kemp, Marjolein van Asselt, Frank Geels, Geert Verbong and Kirsten Molendijk (2000), *Transities & Transitiemanagement: De casus van een Emissiearme Energievoorziening*, Report for Ministry of Environment, Maastricht: International Centre for Integrated Studies.

Sahal, D. (1985), 'Technological guideposts and innovation avenues', *Research Policy*, **14** (2), 61–82.

Sanchez, R. and J.T. Mahoney (1996), 'Modularity, flexibility, and knowledge management in product and organization design', *Strategic Management Journal*, **17**, 63–76.

Saxenian, Annalee (1994), *Regional Advantage: Culture and Competition in Silicon Valley and Route 128*, Cambridge, MA: Harvard University Press.

Schatzberg, E. (1994), 'Ideology and technical choice: the decline of the wooden aeroplane in the United States, 1920–1945', *Technology and Culture*, **35** (1), 34–69.

Schot, Johan W. (1991), 'Maatschappelijke Sturing van Technische Ontwikkeling: Constructief Technology Assessment als Hedendaags Luddisme', PhD thesis, Twente University, Enschede.

Schot, J.W. (1998), 'The usefulness of evolutionary models for explaining innovation. The case of the Netherlands in the nineteenth century', *History of Technology*, **14**, 173–200.

Schot, Johan W. and Adri A. De la Bruheze (2003), 'The mediated design of products, consumption and consumers in the twentieth century', in N. Oudshoorn and T. Pinch (eds), pp. 229–45.

Schot, J., R. Hoogma and B. Elzen (1994), 'Strategies for shifting technological systems: the case of the automobile system', *Futures*, 26, 1060–76.

Schrag, Z.M. (2000),'The bus is young and honest: transportation politics, technical choice, and the motorization of Manhattan surface transit, 1919–1936', *Technology and Culture*, **41**, 51–79.

Schwartz-Cowan, Ruth (1987), 'The consumption junction: a proposal for research strategies in the sociology of technology', in Wiebe E. Bijker, Thomas P. Hughes and Trevor J. Pinch (eds), *The Social Construction of Technological Systems: New Directions in the Sociology and History of Technology*, Cambridge, MA: MIT Press, pp. 261–80.

Scott, W.R. (1995), *Institutions and Organizations*, Thousand Oaks, CA and London, UK: Sage.

Segal, Howard P. (1986), 'The technological utopians', in Joseph J. Corn (ed.), *Imagining Tomorrow: History, Technology and the American Future*, Cambridge, MA: MIT Press, pp. 119–36.

Simon, Herbert A. (1957), *Administrative Behavior: A Study of Decision-Making Processes in Administrative Organization*, 2nd edn, New York: Macmillan.

Simon, H.A. (1962), 'The architecture of complexity: hierarchic systems', *Proceedings of the American Philosophical Society*, **106** (6), 467–82.

Simon, Herbert A. (1973), 'The organization of complex systems', in Pattee (ed.) (1973a), pp. 1–27.

Sloan, Edward W. (1998), 'Private enterprise as public utility: the management of capital in two centuries of shipping business', in Paul C. Van Royen, Lewis R. Fischer and David M. Williams (eds), *Frutta di Mare: Evolution and Revolution in the Maritime World in the 19th and 20th Centuries*, Amsterdam: Batavian Lion International, pp. 89–100.

Smith, Edgar C. (1937), *A Short History of Naval and Marine Engineering*, Cambridge: Cambridge University Press.

Smith, Michael J.T. (1989), *Aircraft Noise*, Cambridge: Cambridge University Press.

Sørensen, Knut H. (2002), 'Social shaping on the move? On the policy relevance of the social shaping of technology perspective', in Knut H. Sørensen and Robin Williams (eds), *Shaping Technology, Guiding Policy: Concepts, Spaces and Tools*, Cheltenham, UK and Northampton, MA, USA: Edward Elgar, pp. 19–35.

Spar, Debora (2001), *Pirates, Prophets and Pioneers: Business and Politics along the Technological Frontier*, London: Random House.

Stankiewicz, Rikard (1992), 'Technology as an autonomous socio-cognitive system', in Hariolf Grupp (ed.), *Dynamics of Science-based Innovation*, Berlin: Springer-Verlag, pp. 19–44.

Stankiewicz, Rikard (2000), 'The concept of "design space", in John Ziman (ed.), *Technological Innovation as an Evolutionary Process*, Cambridge: Cambridge University Press, pp. 243–7.

Staudenmaier, John M. (1989), 'The politics of successful technologies', in Steven Cutliffe and Robert C. Post (eds), *In Context: History and the*

History of Technology: Essays in Honor of Melvin Kranzberg, Bethlehem, PA: Lehigh University Press, pp. 150–71.

Strum, Shirley and Bruno Latour (1999), 'Redefining the social link: from baboons to humans', in Donald MacKenzie and Judy Wacjman (eds), *The Social Shaping of Technology*, Buckingham: Open University Press, 2nd edn, pp. 116–25

Summerton, Jane (ed.) (1994), *Changing Large Technical Systems*, Boulder, CO and Oxford, UK: Westview.

Te Velde, Robin (1997), *Van Technologische Niche tot Dominant Regime: De Verspreiding van de Benzineauto 1859–1906*, Report, Enschede: University of Twente.

Thompson, F.M.L. (1976), 'Nineteenth-century horse sense', *Economic History Review*, **29** (1), 60–81.

Tillinghast Jr., Charles C. (1966), 'The air transport operator and the economic trend', in *'Conjuncture' of Air Transport: The Interaction Between Air Transport and General Technical, Economic and Social Trends*, Paris: Institute of Air Transport (ITA), pp. 1–20.

Todd, Daniel and Jamie Simpson (1986), *The World Aircraft Industry*, London: Croom Helm.

Trist, Eric and K.W. Bamforth (1951), 'Some social and psychological consequences of the long-wall method of coal-getting', *Human Relations*, **4**, 3–38.

Trist, Eric, Fred Emery and Hugh Murray (eds) (1997), *The Social Engagement of Social Science: A Tavistock Anthology, Vol. III: The Socio-Ecological Perspective*, Philadelphia, PA: University of Pennsylvania Press.

Trist, Eric and Hugh Murray (eds) (1990), *The Social Engagement of Social Science: A Tavistock Anthology, Vol. I: The Social Psychological Perspective*, Philadelphia, PA: University of Pennsylvania Press.

Trist, Eric and Hugh Murray (eds) (1993), *The Social Engagement of Social Science: A Tavistock Anthology, Vol. II: The Socio-technical Perspective*, Philadelphia, PA: University of Pennsylvania Press.

Tushman, M. and P. Anderson (1986), 'Technological discontinuities and organization environments', *Administrative Science Quarterly*, **31**, 393–465.

Tushman, M.L. and J.P. Murmann (1998), 'Dominant designs, technology cycles, and organizational outcomes', *Research in Organizational Behaviour*, **20**, 231–66.

Tushman, Michael L. and Lori Rosenkopf (1992), 'Organizational determinants of technological change: towards a sociology of technical evolution', in Larry L. Cummings and B.M. Staw (eds), *Research in Organizational Behavior*, 14, Greenwich, CT: JAI Press, pp. 311–47.

Tylecote, Andrew (1993), *The Long Wave in the World Economy: The Current Crisis in Historical Perspective*, London and New York: Routledge.

Ulrich, Kurt (1998), *Monarchs of the Sea: The Great Ocean Liners*, London: Tauris Parke.

Unruh, G.C. (2000), 'Understanding carbon lock-in', *Energy Policy*, **28** (2000), 817–30.

Utterback, James M. (1994), *Mastering the Dynamics of Innovation*, Boston, MA: Harvard Business School Press.

Utterback, J.M. and W.J. Abernathy (1975), 'A dynamic model of process and product innovation', *OMEGA*, **3** (6), 639–56.

Van de Poel, Ibo (1998), '*Changing technologies. A comparative study of eight processes of transformation of technological regimes*', PhD thesis, Twente University, Enschede.

Van de Poel, I. (2000), 'On the role of outsiders in technical development', *Technology Analysis and Strategic Management*, **12**, 383–97.

Van de Ven, Andrew H., Harold L. Angle and Marshall Scott Pode (eds) (1989), *Research on the Management of Innovation: The Minnesota Studies*, New York: Harper & Row.

Van de Ven, Andrew H. and Raghu Garud (1994), 'The coevolution of technical and institutional events in the development of an innovation', in Baum and Singh (eds), pp. 425–43.

Van den Ende, J. and R. Kemp (1999), 'Technological transformations in history: how the computer regime grew out of existing computing regimes', *Research Policy*, **28**, 833–51.

Van der Linden, F. Robert (1995), 'Progressives and the post office: Walter Folger Brown and the creation of United States air transportation', in William F. Trimble (ed.), *From Airships to Airbus: The History of Civil and Commercial Aviation, Volume 2: Pioneers and Operations*, Washington, DC and London, UK: Smithsonian Institution Press, pp. 245–60.

Van Dijck, José (1998), *Imagenation: Popular Images of Genetics*, New York: New York University Press.

Van Driel, H. and J. Schot (2001), 'Regime-transformatie in de Rotterdamse graanoverslag', *NEHA-jaarboek voor Economische, Bedrijfs- en Techniekgeschiedenis*, **64**, 286–318.

Van Lente, Harro (1993), '*Promising technology: the dynamics of expectations in technological development*', PhD thesis, Twente University, Delft: Eburon.

Van Lente, Harro (1995), 'The rise of technology studies in the Netherlands', in Tarja Cronberg and Knut H. Sørensen (eds), *Similar Concerns, Different Styles? Technology Studies in Western Europe*, Luxembourg: Office for Official Publications of the European Communities, pp. 287–325.

Van Lente, Harro and Arie Rip (1998), 'Expectations in technological developments: an example of prospective structures to be filled in by agency', in Cornelis Disco and Barend J.R. van der Meulen (eds), *Getting New Technologies Together*, Berlin, New York: Walter de Gruyter, pp. 195–220.

Van Oosten, F.C. (1972), *Schepen onder Stoom: De Geboorte van het Stoomschip*, Bussum, Holland: Unieboek, B.V.

Verbong, Geert P.J. and Peter van Overbeeke (1994), 'De compoundmachine', in Harry W. Lintsen (ed.), *Geschiedenis van de techniek in Nederland: De wording van een moderne samenleving, 1800–1890, deel V*, Zutphen: Walburg Pers, pp. 222–41.

Ville, Simon P. (1990), *Transport and the Development of the European Economy: 1750–1918*, Basingstoke: Macmillan.

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

Vincenti, W.G. (1994), 'The retractable airplane landing gear and the Northrop "anomaly": variation-selection and the shaping of technology', *Technology and Culture*, **35**, 1–33.

Vincenti, W.G. (1995), 'The technical shaping of technology: real-world constraints and technical logic in Edison's electric lighting system', *Social Studies of Science*, **25** (3), 553–74.

Volti, R. (1996), 'A century of automobility', *Technology and Culture*, **37**, 663–85.

Von Hippel, Eric (1988), *The Sources of Innovation*, Oxford: Oxford University Press.

Von Meier, Alexandra (1994), 'Integrating supple technologies into utility power systems: possibilities for reconfiguration', in Summerton (ed.), pp. 211–30.

VROM (2001), *Een wereld en een wil, werken aan duurzaamheid, Nationaal Milieubeleidsplan 4*, Dutch Environmental Policy Plan 4, The Hague: Ministry of VROM.

Wakefield, Ernest H. (1994), *History of the Electric Automobile: Battery-only Powered Cars*, Warrendale, PA: Society of Automotive Engineers.

Walker, W. (2000), 'Entrapment in large technology systems: institutional commitments and power relations', *Research Policy*, **29**, 833–46.

Ward, W.H. (1967), 'The sailing ship effect', *Bulletin of the Institute of Physics and Physical Society*, **18**, 169.

Weber, Mathias (1997), 'Innovation Diffusion and Political Control of Energy Technologies: A Comparison of Combined Heat and Power Generation in the UK and Germany', PhD thesis, Institut für Sozialforschung, University of Stuttgart, Stuttgart, Germany.

Weterings, R., J. Kuijper, E. Smeets, G.J. Annokkée and B. Minne (1997), *81 mogelijkheden Technologie voor duurzame ontwikkeling. Eindrapport van de milieugerichte technologieverkenning*, Report, The Hague: Ministry of VROM.

Williams, R. and D. Edge (1996), 'The social shaping of technology', *Research Policy*, **25** (6), 865–99.

Willis, Carol (1986), 'Skyscraper utopias: visionary urbanism in the 1920s', in Joseph J. Corn (ed.), *Imagining Tomorrow: History, Technology and the American Future*, Cambridge, MA: MIT Press, pp. 164–87.

Wimmers, Henk J. (1998), 'Evolution and revolution in the development of shipbuilding materials', in Paul C. Van Royen, Lewis R. Fischer and David M. Williams (eds), *Frutta di Mare: Evolution and Revolution in the Maritime World in the 19th and 20th Centuries*, Amsterdam: Batavian Lion International, pp. 123–32.

Winner, L. (1980), 'Do artefacts have politics?', *Daedalus*, **109** (1), 121–36.

Winskel, M. (2002), 'When systems are overthrown: the "dash for gas" in the British electricity supply industry', *Social Studies of Science*, **32** (4), 563–98.

Yago, Glenn (1984), *The Decline of Transit: Urban Transportation in German and U.S. Cities, 1900–1970*, Cambridge, UK and New York: Cambridge University Press.

Yin, Robert K. (1994), *Case Study Research: Design and Methods*, 2nd edn, Thousand Oaks, CA: Sage.

Zeitlin, J. (1995), 'Flexibility and mass production at war: aircraft manufacture in Britain, the United States, and Germany, 1939–1945', *Technology and Culture*, **36**, 46–79.

Index