The Hidden Dynamics of Path Dependence

Institutions and Organizations

Edited by Jörg Sydow
and Georg Schreyögg

The Hidden Dynamics of Path Dependence

This page intentionally left blank

The Hidden Dynamics of Path Dependence

Institutions and Organizations

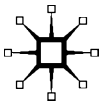
Edited by

Georg Schreyögg
Professor of Organization and Leadership, Freie Universität Berlin

and

Jörg Sydow
Professor of Management, Freie Universität Berlin

palgrave
macmillan



Softcover reprint of the hardcover 1st edition 2010 978-0-230-22081-2

First published 2010 by
PALGRAVE MACMILLAN

Palgrave Macmillan in the UK is an imprint of Macmillan Publishers Limited, registered in England, company number 785998, of Houndmills, Basingstoke, Hampshire RG21 6XS.

Palgrave Macmillan in the US is a division of St Martin's Press LLC, 175 Fifth Avenue, New York, NY 10010.

Palgrave Macmillan is the global academic imprint of the above companies and has companies and representatives throughout the world.

ISBN 978-1-349-30630-5 ISBN 978-0-230-27407-5 (eBook)
DOI 10.1057/9780230274075

This book is printed on paper suitable for recycling and made from fully managed and sustained forest sources. Logging, pulping and manufacturing processes are expected to conform to the environmental regulations of the country of origin.

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

A catalog record for this book is available from the Library of Congress.

10 9 8 7 6 5 4 3 2 1
19 18 17 16 15 14 13 12 11 10

Contents

Tables and Figures

Tables

Figures

Contributors

Arjen Boin is an associate professor in the Public Administration Institute at Louisiana State University. He publishes on topics of crisis management, administrative leadership and institutional development. He is the US editor of the journal Public Administration.

Professor **Dan (Danny) Breznitz** (Georgia Institute of Technology Sam Nunn School of International Affairs & The School of Public Policy, PhD MIT) has extensive experience in conducting comparative in-depth research of Rapid-Innovation-Based Industries and their globalization. Dr. Breznitz's latest book, *Innovation and the State: Political Choice and Strategies for Growth in Israel, Taiwan, and Ireland (Yale University Press)*, won the 2008 Don K. Price for best book on science and technology given by APSA and was a finalist for the 2007 best book of the year award in political science by *ForeWord* Magazine. He is in the midst of finishing another book manuscript (co-authored with Michael Murphree) *The Run of the Red Queen: Government, Innovation, Globalization, and Economic Growth in China*. Breznitz is one of five young North American scholars to be selected as a 2008 Industry Study Fellow of the Sloan Foundation.

Olof Brunninge is Assistant Professor in Business Administration at Jönköping International Business School (JIBS) in Sweden. He defended his doctoral dissertation 'Organizational Identity and the Strategy Process' in 2005. His main research and teaching interests are in the fields of strategic management, organizational identity, family businesses, and the role of history in strategic management.

Robert A. Burgelman is the Edmund W. Littlefield Professor of Management at the Stanford University Graduate School of Business. He holds a licentiate degree in Applied Economics from Antwerp University, and an MA in Sociology and a Ph.D. in Management of Organizations from Columbia University. In 2003 he received an honorary doctorate from the Copenhagen Business School for his contributions to the study of corporate innovation.

Dr. Michael W. Conaway graduated from Washington State University College of Veterinary Medicine with honours in 1978. He grew up in Scottsdale, Arizona with horses and decided to become an equine

veterinarian at an early age. After graduation, he moved to Tucson and started his own practice. His special interests include equine reproduction & infertility problems, lameness, and laminitis.

Stephan von Cramon-Taubadel has a Chair for Agricultural Policy at the Georg-August-University of Göttingen. He received his BSc and MSc degrees at McGill University (Montréal) and the University of Manitoba, respectively, before completing his PhD and Habilitation at the Christian-Albrechts-University in Kiel. His specific research interests include quantitative methods of price transmission and market integration analysis, and the transition of agriculture in the Former Soviet Union.

Wilfred Dolfsma is professor of Innovation and Strategy at the University of Groningen, School of Economics and Business. His research focuses on the role, advantages and disadvantages of cooperation between parties in processes of innovation. Recent work includes *Knowledge Economies* (Routledge, 2008) and *Institutions, Communication and Values* (Palgrave, 2009), and has appeared in such journals as Research Policy and the Journal of Business Ethics. He is corresponding editor of the Review of Social Economy.

Hugo van Driel is assistant professor of Business History at the Rotterdam School of Management, Erasmus University. He is the co-author of recent publications on changing backgrounds of top managers (in: Harm Schröter, *The European Enterprise*, Springer, 2008), path dependence of cooperative forms in port business (*Business History Review*, 2007), and the 'industrial logic' of middle management intensity between 1900 and 1950 (*Business History*, 2007).

Dr. Myles Gartland spends his research and teaching time in the areas of competitive analysis & strategy formulation as well as issues related to globalization. He is interested in issues such as industry evolution and boundaries of the firm; with specific interest in the health care industry. He has published in numerous management and economics journals.

Dr. Gerhard Fuchs works at the University of Stuttgart (Department of Social Sciences). Before that he acted as a Deputy Director of the Department Technology, Organization, Work at the Centre for Technology Assessment in Baden-Württemberg, Stuttgart (Germany). Research interests: Regional Innovation Systems, Innovation Research.

Sebastian Hess is a researcher at the Department of Economics, Swedish University of Agricultural Sciences (SLU), Uppsala, Sweden. He studied Agricultural Economics at the University of California (Davis) and in

Göttingen, where he also obtained his PhD. His research is focused on the analysis of agricultural policies and rural labour markets.

Daniela Kleinschmit is assistant professor at the Swedish University of Agricultural Sciences (SLU) in Uppsala, Sweden. Her research interests include resource governance and democracy theory, focusing on discourse and legitimacy.

Dr. Kenneth W. Koput has been a professor in the Eller College of Management at the University of Arizona since 1992. Dr. Koput received his doctorate from The University of California, Berkeley, specializing in Organizational Behaviour and Industrial Relations. Since then, he has published papers on industry structure and innovation in several leading academic journals and has won a number of national and international awards for his scholarship.

Sanneke Kuipers is assistant professor at Leiden University's Department of Public Administration, where she heads the university's Crisis Research Center. In addition, she is a senior advisor at Crisisplan BV, an international consultancy firm specialized in crisis management. Her dissertation was titled The Crisis Imperative: Crisis Rhetoric and Welfare State Reform in Belgium and the Netherlands in the Early 1990s (Amsterdam University Press, 2006) and focused on how crisis rhetoric can be used to change path dependent public policies. Her research was awarded with two national prizes. She also published on policy evaluation, accountability processes after crisis, crisis-induced reform, EU crisis management, and EU decision making.

Gérard Marty is a qualified teacher, holder of the 'agrégation' (highest competitive examination for teachers in France) in Economics and Management Science. His research, realized at the Laboratory of Forestry Economics (UMR ENGREF-INRA) and at the Sorbonne Sociological Studies Centre (University of Paris-Sorbonne, Paris 4) is focused on the analysis of the social construction of markets.

Leif Melin is Professor of Strategy and Organization at Jönköping International Business School (JIBS) in Sweden. He has been the founding Director of CeFEO – Center for Family Enterprise and Ownership at JIBS since 2005. Leif Melin's research interests include several areas related to strategizing and strategic change. He has been active in the development of a new, emerging perspective within strategic management; the strategy-as-practice-perspective.

Kim van Nieuwaal studied at the National and Kapodastrian University of Athens, Leiden University and the Erasmus University of Rotterdam,

getting his master's degree in Public Administration from the latter in 2000. Subsequently, he worked as IT-specialist in Sydney, Australia. Since 2002 he has been affiliated with VU University of Amsterdam, teaching courses on, amongst others, Public Administration and Strategic Management and conducting research on the firm-government interaction. His dissertation, called 'The Institutional Survival Path of the Firm', is due late 2009. Kim currently works as scientific coordinator for the Dutch national research programme Knowledge for Climate and is affiliated with the Faculty of Geosciences of Utrecht University.

Dr. David Olson is an alumnus of the University of Arizona, having received his Ph.D. in Management in 1998. Dr. Olson now teaches Business Strategy and Entrepreneurship as an Associate Professor at California State University at Bakersfield. His research interests include resource dependence and market entry.

Kaya Özkaracalar is the head of the Film and Television Department of Bahcesehir University, Istanbul, Turkey, where he teaches World Cinema and Genre Cinema. He has contributed numerous articles including film reviews to Turkish daily newspapers and monthly film journals as well as editing the quarterly *Geceyarısı Sineması*. He is the author of books titled *Gotik* (L-M, 2006) and *Geceyarısı Filmleri* (+1 Kitap, 2007).

Özlem Öz received her PhD at the London School of Economics and Political Science (LSE). She now works at the Department of Management, Bogazici University, Istanbul, Turkey, where she teaches courses on organisation theory. She is the author of the books titled *The Competitive Advantage of Nations: The Case of Turkey* (Ashgate, 1999) and *Clusters and Competitive Advantage: The Turkish Experience* (Palgrave MacMillan, 2004).

Georg Schreyögg is Professor of Organization and Leadership at Freie Universität Berlin. His current research interests include organizational change, organizational capabilities and path dependence. He received his doctorate in business administration from the Universität Erlangen-Nürnberg. He was International Fellow of AIM (Advanced Institute of Management Research, UK) and Elected Fellow at DFG (Deutsche Forschungsgemeinschaft). In 2005/06 he served as President of the German Association of University Teachers for Business Administration (VHB). He has been a member of the editorial board of several international journals and is co-editor of *Managementforschung*. He has published eight books, nine edited volumes and over 150 contributions, which have appeared in journals such as *Academy of Management Review*,

Strategic Management Journal, Organization Science, Journal of Business Ethics, Organization Studies, Organization, and in numerous other journals and volumes.

Dr. Jan Siedentopp received his PhD from the School of Business & Economics at the Freie Universität Berlin in 2008, where he was a member of the Doctoral Program 'Research on Organizational Paths'. His empirical research projects in the field of strategic management focuses on public affairs management, path management, and resource-based theories.

Albrecht Söllner is the Professor for International Management at Europa-Universität Viadrina at Frankfurt (Oder), Germany, and a visiting scholar at Haas School of Business, University of California Berkeley. He completed a PhD (Dr. rer. pol.) at Freie Universität Berlin in 1993 and a 'Habilitation' (venia legendi) at Humboldt-Universität zu Berlin. Before joining the Viadrina, he held a professorship in Business Administration at Westfälische Wilhelms-Universtity, Münster, Germany. Albrecht Söllner keeps good relations to a variety of business firms and served as a member of the board of directors of the biotech-firm Capsulution Nano Science.

Dr. Martin Stack is an Associate Professor of Management at Rockhurst University. His primary research areas are in the health care and the brewing industries. He is particularly interested in examining how these traditionally national markets are increasingly internationalizing. He has also done extensive research in the related areas of path dependence, path creation, and behavioural lock-in. He has published a number of book chapters and articles in journals such as Business History Review, Business Horizons, Journal of Management History, and the Journal of Economic Issues.

Jörg Sydow has earned his doctorate and habilitation in business administration at the Freie Universität Berlin where he currently is a Professor of Management at the School of Business & Economics. Before he held a chair for planning and organization at the University of Wuppertal, Germany, and was a visiting professor at Bentley College, the Universities in Innsbruck and Vienna, the University of Arizona, and the University of Strathclyde. From 2004 until 2007 he was an international visiting fellow of the Advanced Institute of Management Research (AIM), London. He is a founding co-editor of two leading German journals, *Managementforschung* and *Industrielle Beziehungen – The German Journal of Industrial Relations*, and a member of the editorial

boards of *Organization Studies*, *Organization Science* and *The Scandinavian Journal of Management*.

Ludwig Theuvsen has a Chair for Management in Agribusiness at the Georg-August-University of Göttingen, Germany. He holds a Diploma in Business Administration from the University of Aachen and a Ph.D. from the University of Cologne. His main research interests include organization of food supply chains and strategic, human resource and quality management in agribusiness firms.

Ulrike Zschache M.A., born 1980 in Dresden (Germany), studied Cultural Sciences, Sociology, Political Sciences and Journalism at Universities in Leipzig (Germany) and Rome (Italy); Thesis on 'Europe in the Media Discourse: The Debate on the Accession of Turkey to the European Union in German Newspapers'; since March 2008 research assistant at Georg-August-University of Göttingen, Department of Agricultural Economics and Rural Development.

Part I
Theoretical Foundations

1
Understanding Institutional and Organizational Path Dependencies*

Georg Schreyögg and Jörg Sydow

1.1 Introduction

The notion of path dependence has become so popular that it is almost commonplace to describe the development of institutions and organizations as being path dependent.[1] This is as true for strategic management and organization theory as it is for economic geography and institutional analysis of society. With its increasing popularity, however, the notion of path dependence has progressively lost a specific meaning. Most often it is merely used as a metaphor accentuating that history matters when explaining cultural artefacts. This is all the more regrettable as problems of path dependence seem to trouble an increasing number of institutions and organizations. No doubt explanations of persistence and rigidity can profit a lot from a re-sharpened concept of path dependence building on the ideas as originally put forward by Paul David (1975, 1985, 1986) and W. Brian Arthur (1989, 1994) in their economic investigation of the diffusion of technologies, the QWERTY keyboard layout being their most prominent example. Some institutional and evolutionary economists have taken up this analytical understanding of path dependence and extended its usage to the institutional area (e.g., North, 1990; Witt, 1997). The same is true for studies in historical institutionalism and comparative politics (e.g., Pierson, 2000), but the bulk of research in these areas, as well as in the field of organizational analysis, refers to the notion of path dependence only in a rather loose way.

As a result of this tendency, the analytical power of (concepts from) the theory of path dependence, by and large, remains underutilized. It

is the aim of this edited volume – and consequently of this introduction to the book – to contribute to a better understanding of institutional persistencies and organizational inertia by referring to the theory of path dependence in a more rigorous manner and to unpack the often hidden dynamics underlying path-dependent processes. In this introductory chapter we will provide a sharpened platform for understanding institutional and organizational path dependence which at the same time serves as an umbrella for the contributions collected for this volume. Many of the contributions have been presented at the international conference 'Studying Path Dependencies of Businesses, Institutions and Technologies' organized by the 'Research on Organizational Paths' center at the School of Business & Economics of Freie Universität Berlin, Germany, but also at a workshop of the European Institute of Advanced Studies of Management (EIASM), at colloquia of the European Group of Organization Studies (EGOS), and at annual meetings of the Academy of Management (AOM). Finally, this introductory chapter will give an overview of the contributions to this volume.

1.2 Organizational path dependence: theoretical framework

The basic point of any path-dependence argument provided so far stresses the importance of past events for future action or, more precisely, of foregoing decisions for current and future decision making. Hence, decisions are conceived as historically conditioned: 'bygones are rarely bygones' (Teece, Pisano and Shuen, 1997: 522). Although this insight has certainly advanced the building of a time-sensitive understanding of institutional and organizational phenomena and helped to overcome the a-historical and unbounded view of rational choice thought, it is too vague a perspective. Path dependence means more than a mere 'past-dependence' (Antonelli, 1999) as such an all-embracing understanding of path dependence does not enable us to better understand the specific dynamics of locked institutional and organizational processes and their underlying forces.

In a first attempt to develop a more specific understanding we can say that path dependence implies at the very least a sequence of events narrowing the scope of action eventually resulting in a state of persistence or inertia. Valuable insights in the dynamic nature of entrapping or locking processes have been provided by studies from evolutionary economics (David 1985, 1986, 1993, 2007; Arthur, 1989, 1994; Dosi, 1982, 1997). In those accounts, decisions taken in the past can increasingly restrain future choices. The decision pattern used repeatedly in

the past may even amount to an imperative for the future course of action so that ultimately no further choice is left. In order to explain the dynamics of these processes, self-reinforcing processes have been identified as major drivers which are likely to accumulate in a specific path of action. The inherent self-reinforcing dynamics – being driven typically by the 'externalities' arising from the actions of other agents, or strongly routinized ('learned') individual behaviors – is likely to pass quickly beyond the control of any individual actor; the regime takes the lead and potentially renders the process into a non-reversible state of local equilibrium or 'lock-in' (David, 1985, 1993).

This brief illustration also highlights that the process of shaping a path cannot be characterized by the same structural properties at any point in time in the process. The initial situation differs completely from the final situation. Therefore, we suggest subdividing the whole process into different stages governed by basically different regimes. We propose distinguishing three phases in the process of bringing about path dependence (cf. Sydow et al., 2009):

Phase I – the Preformation Phase – can be characterized by a broad scope of action, where choices taken cannot be predicted by prior events or initial conditions (Mahoney, 2000: 511). Up to a certain degree the first phase is, however, also influenced by the past. Thus, Phase I should conceptually build on historically framed and 'imprinted' contingency and not on completely unrestricted choice as rational choice models do. Once a decision is made – on whatever basis – this choice may turn out to be a 'small event.' If so, this 'innocent' decision sets into motion self-reinforcing processes. The shadow in Phase I in Figure 1.1 is supposed to indicate this thought. The moment of entering into the dynamics of a self-reinforcing process can be conceived of as a 'critical juncture' (Collier and Collier, 1991).

In Phase II – the Formation Phase – the range of options increasingly narrows and it becomes progressively difficult to reverse the initial choice, i.e. a path is evolving. In this phase of path-building a new regime takes the lead: the dynamics of self-reinforcing processes, conceptualized by Arthur (1989, 1994) as increasing returns. On a general level, the notion of increasing returns highlights positive feedback processes in which the increase of a particular variable leads to a further increase of this very variable. The notions of increasing returns or positive feedback indicate self-reinforcing processes with growing benefits which are likely to culminate in a patterned dynamic. Eventually, a dominant institutional or organizational solution emerges. The other side of the coin is that the probability of the direction of evolution exhibited by the macro-process being reversed is tending to decrease with each successive step of its

unfolding, and in that sense the system is becoming committed to only one or a small subset of the originally possible outcomes. Decision processes in Phase II are, however, still contingent, i.e. essentially constrained choices are still possible. As David (1985, 2007) puts it, the processes are 'non-ergodic,' i.e., their motion is neither pre-determined nor random but their evolvement proceeds under the influence of its own history, which in effects 'selects' one or another among the many outcomes that potentially might have been attained.

The empirical studies that have provided ample evidence for the existence of self-reinforcing and non-reversible dynamics have been primarily conducted in the field of technology development and diffusion. In economic studies (Arthur, 1994; Cowan, 1990; Katz and Shapiro, 1985, North 1990) so far, basically six types of self-reinforcing dynamics can be identified: economies of scale, (direct and indirect) network externalities, learning effects, adaptive expectations, coordination effects, and complementarities. Examining the economic effects in more detail reveals, however, that they develop under special conditions only, namely if the decision to reproduce a particular option is suggested by a utility calculus. Without excluding this case, it seems too restrictive a starting point for explaining institutional and organizational paths (see also Ortmann, 1995; Crouch and Farrell, 2004; Eden, 2004). As is well known from a long-standing research tradition, there are other potentially salient self-reinforcing patterns in organizations, for instance, based on emotional reactions, cognitive biases, political processes, and so forth (e.g., Huff and Huff, 2000). These dimensions have to be included in order to understand the dynamics of institutional and organizational paths. Our suggestion, therefore, is to broaden the scope of self-reinforcing effects by including all kinds of positive feedback cycles. In this broader feedback category, increasing returns represent a specific form of self-reinforcement – among other forms of positive feedback. The institutional and organizational studies mentioned above also urge to adopt a broader perspective beyond individual decision-making, which accounts for the social setting in which the positive feedback processes are embedded. It is the broader context (the sedimented institutions, the hidden assumptions of the organization, the organizational culture, the status and role system, etc.) which informs decision makers and provides, indirectly and inadvertently, the drivers of self-reinforcing loops.

Take, for instance, learning effects. The original version holds that the more often an operation is performed, the more efficiency will be gained when operating subsequent iterations. The operations become more skillful (faster, less errors, etc.), which, in turn, means decreasing average cost

per unit of output. And, the more attractive the chosen solution becomes due to accumulated skills and decreasing cost, the less attractive is switching to other learning sites (where these actors would have to start from scratch). Only sticking with the once chosen solution promises (increasing) returns. However, there is not only the positive effect of increasing returns from learning. Rather, the literature on organizational learning points to the fact that a focus on exploitative learning driven by increasing returns may gradually drive out explorative learning (March 1991, 2006). For various reasons (prevailing institutional rules, organizational culture, reward system, etc.) the motivation to improve everyday practices is likely to get more acceptance or legitimacy, whereas the motivation to look for fresh alternatives and to critically examine well-established institutional or organizational practices shrinks step by step. This myopia or preference for exploitation learning eventually builds another path along the familiar practices. A related effect has been highlighted by Miller's (1993) 'Architecture of Simplicity': an organization develops a successful set of strength and tends to focus all learning abilities on refining this success. It amplifies this strength while neglecting other opportunities. Ultimately, 'it turns into a monolithic narrowly focused version of its former self, converting a formula for success into a path toward failure' (Miller, 1993: 116). Basically, these and the other effects resulting from self-reinforcing dynamics refer to both push and pull explanations of path dependence. It should be stressed that the effects are not fully separate; they rather often occur jointly and overlap.

The transition from this second phase to Phase III – the Lock-in Phase – can be characterized through a further constriction which is likely to (but will not necessarily) lead into a lock-in, i.e., the dominant pattern gets fixed and gains a quasi-deterministic character. Further processing is fully bound to a path. One particular choice or action pattern has become the predominant mode; flexibility has been lost. Even new entrants into this field of action have to adopt it. The problem of a lock-in becomes particularly obvious in cases where a more efficient alternative appears but a switch is no longer possible, and the result is actual or potential inefficiency.

Considering institutional and organizational paths, the context seems to require a somewhat modified conception of lock-in. Due to its social character, institutional and organizational processes are more complex and ambiguous in nature. They are not likely to amount to a concrete monopolistic solution, which excludes any further choices. Rather, they are likely to bring about a specific action pattern which gets deeply embedded in practice and replicated across various situations.

Fixed recursive patterns of action, however, cannot be equated with determinism, not forgetting the power of knowledgeable agents to act otherwise (Giddens, 1984). For these reasons, we should refrain from reifying paths and attributing them an objective quality. Instead of a fully determined lock-in, no matter whether of a cognitive, normative or resource-based kind, a theory of institutional or organizational paths is well advised to rather conceptualize the final stage of a path-dependent process in a less restrictive way leaving some scope for variation (similarly Thelen, 1999; Pierson, 2000; Eden, 2004; Martin and Sunley, 2006). By implication, it seems more adequate to conceive of the lock-in state not in terms of total rigidity but rather as a matter of degree accounting for variance in actual practicing of the path. A corridor may serve best to illustrate this reasoning; the shadow in Phase III in Figure 1.1 is designed to indicate this corridor.

In sum, the proposed theory – as shown in Figure 1.1 – conceptualizes an institutional or organizational path as a social process that is initially, in Phase I, shaped by contingent choices leading to a critical juncture which is triggered by a small or bigger event. Phase II, the very Path Formation Phase, is governed by a regime of positive, self-reinforcing feedback loops, constituting and setting up a specific pattern of actions (social practices), which gains progressively predominance against alternative solutions to an extent that, at least potentially, leads in Phase III into a cognitive, normative or resource-based lock-in, in terms of a pattern underlying actions and/or practices.

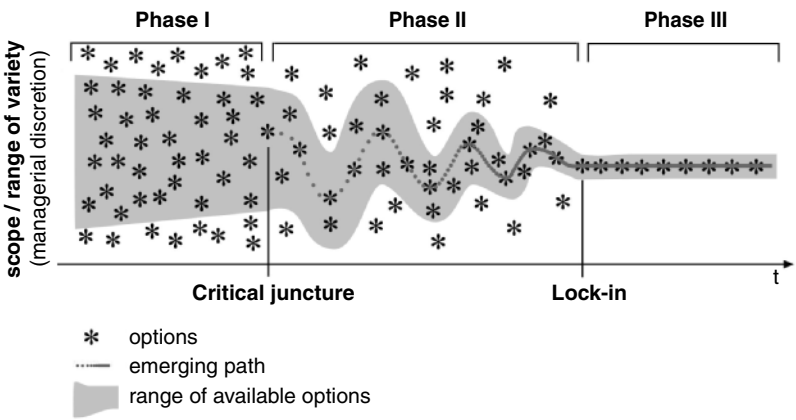


Figure 1.1 The constitution of a path
Source: (Sydow et al., 2009, p. 692).

1.3 Empirical studies of institutional and organizational path dependence

While an uncountable number of studies have investigated the persistence of institutions and/or the rigidity of organizations, only relatively few have adopted an analytical approach based upon (concepts from) the theory of path dependence as originally developed by David and Arthur and later modified by social scientists such as Kathy Thelen (1999), Paul Pierson (2000) and Jim Mahoney (2000). All explanations of institutional and organizational persistence provided so far, point to critical features and have their merits. What they are missing, however, is a theoretical clarification of the process of becoming persistent. What are the underlying dynamics and why do people not resist institutional rigidification? In which cases do organizations get locked into their behavioural patterns and in which not? Are there major driving forces? In the main, the contributions to this volume follow such an analytical approach and sometimes contrast it with other forms of institutional or organizational rigidification, such as imprinting (e.g., Johnson, 2007), for instance.

The volume is organized along the following rationale: After this introduction it starts with a contribution by Dan Breznitz on contradictions and pitfalls of the path-dependence debate. His chapter concludes by providing a revised framework designed to avoid major shortcomings.

The second part of the volume is on Path Dependence in Public and Private Organizations. The first paper by Hugo van Driel and Wilfred Dolfsma analyzes the Toyota production system from an imprinting point of view. Meta-routines structure the innovation process, thereby binding innovation initiatives (inadvertently) to routines. Path dependence may be the longer-term result. The second paper by Sanneke Kuipers and Arjen Boin reports from studies on public bureaucracies, the institutionalization processes and their tendency to become path dependent.

The third part of the volume focuses on Strategic Path Dependence. The opening contribution by Jan Siedentopp and Albrecht Söllner sheds light on corporate political activity and their unintended consequences in terms of narrowing the scope of action and continuing recursive reproduction. The other strategy paper by Olof Brunninge and Leif Melin deals with multinational enterprises and demonstrates strategic conservatism (path dependence) even in times of deliberate change.

The fourth part of the volume highlights Path-Dependent Industry Development. It includes five papers. The first study, authored by Martin Stack and Myles Gartland, focuses on offshoring of physician services and demonstrates – paradoxically enough – how a modularization

approach eventually led to path dependence. The second industry paper by Ken Koput, Michael Conaway and David Olson provides a striking example of puzzling persistence in the horse-shoe industry. Despite relatively small switching costs, no switch occurs to non-traditional horse shoes. The authors explain this by path dependence. Gerard Marty analyzes in his contribution to this part of the volume a very special public institution, the French timber auctions. He demonstrates quite convincingly that this century-long institution survived despite changed environmental conditions and despite serious doubts in his ongoing efficiency – another case of path dependence. Özlem Öz and Kaya Ozkaracalar in their paper shed light on the Turkish film industry and the development of a regional cluster which has become surprisingly inert despite the character of the region as a 'project ecology' (Grabher, 2005). Finally, Gerhard Fuchs also focuses on a regional cluster in his analysis, in his case on Baden-Württemberg in the southwest of Germany and its astonishing development between persistence and adaptability.

The fifth part of the volume focuses on Path Dependence of Public Policy and Institutions and comprises two papers. In the first one, Sebastian Hess, Daniela Kleinschmit, Ludwig Theuvsen, Stephan von Cramon-Taubadel and Ulrike Zschache set out to explain the dependence of the German agrarian industry on seasonal farm workers and the difficulties of escaping it. The authors see the major cause in a path-dependent policy making of the German government. In the second, Kim van Nieuwaal analyzes the debate on institutional path dependence taking the example of the Dutch small-fields policy on gas depletion in the Wadden Sea. Focusing on controversies the question he poses – and answers – is why this policy has remained intact for more than three decades, despite good arguments against it.

The final part of the volume offers Perspectives on a Theory of Path Dependence as it is and should be applied to the understanding of institutions and organizations that are or become persistent or inert. Robert Burgelman reports on insights from his outstanding research on strategy development and co-evolutionary lock-ins. He suggests sticking closely to the case and to refrain from imposing hastily theoretical conceptions on real-life cases.

Notes

* In preparing the volume we are indebted to Paul David for comments on an earlier version of this introduction, to Elke Schüßler for assisting the editorial

process, and to Irmgard Hoemke for copy editing. We also received a great deal of support from the 'Research on Organizational Paths' center at Freie Universität, and the colleagues and the doctoral students at this center. Last but not least we gratefully acknowledge the generous funding of the center by the German Research Foundation (DFG).

1. A quick look at the SSCI (search: 'path dependence' or 'path dependency') proves a remarkable growth of path-dependence publications and citations starting from 1996.

References

Antonelli, C. (1999) 'The economics of path-dependence in industrial organization', *International Journal of Industrial Organization*, vol. 15, pp. 643–675.

Arthur, W.B. (1989) 'Competing technologies, increasing returns, and lock-in by historical events', *Economic Journal,* vol. 99, pp. 116–131.

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

Collier, R.B. and D. Collier (1991) *Shaping the Political Arena: Critical Junctures, the Labor Movement, and Regime Dynamics in Latin America*. Princeton, PN: Princeton University Press.

Cowan, R. (1990) 'Nuclear power reactors: A study in technological lock-in', *The Economic Journal*, vol. 106, pp. 541–567.

Crouch, C. and H. Farrell (2004) 'Breaking the path of institutional development? Alternatives to the new determinism', *Rationality and Society*, vol. 16 (1), pp. 5–43.

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

David, P.A. (1985) 'Clio and the economics of QWERTY', *American Economic Review*, vol. 75 (2), pp. 332–337.

David, P.A. (1986) 'Understanding the economics of QWERTY: The necessity of history', in W.N. Parker (ed.) *Economic History and the Modern Economist*. Oxford: Blackwell, pp. 30–49.

David, P.A. (1993) 'Path-dependence and predictability in dynamic systems with local network externalities: A paradigm for historical economics', in D. Foray and C. Freeman (eds) *Technology and the Wealth of Nations*. London: Pinter, pp. 208–231.

David, P.A. (2007) 'Path dependence and its critiques its critics and the quest for historical economics', in G.M. Hodgsons (ed.) *The Evolution of Economic Institutions*. Cheltenham: Edward Elgar, pp. 120–142.

Dosi, G. (1982) 'Technological paradigms and technological trajectories: A suggested interpretation of the determinants and directions of technical change', *Economic Journal*, vol. 107, pp. 1530–1547.

Dosi, G. (1997) 'Opportunities, incentives and the collective patterns of technological change', *The Economic Journal*, vol. 107 (September), pp. 1530–1547.

Eden, L. (2004) *Whole World on Fire: Organizations, Knowledge, and Nuclear Weapons Devastation*. Ithaca, NY: Cornell University Press.

Fuchs, G. and P. Shapira (eds) (2005) *Rethinking Regional Innovation and Change*. Berlin: Springer.

Giddens, A. (1984) *The Constitution of Society*. Cambridge: Polity Press.
Grabher, G. (2005) 'Switching ties, recombining teams: Avoiding lock-in through project organization?' in G. Fuchs and P. Shapira (eds) *Rethinking Regional Innovation and Change*. Berlin: Springer, pp. 63–83.
Huff, A.S. and J.O. Huff (2000) *When Firms Change Direction*. Oxford: Oxford University Press.
Johnson, V. (2007) 'What is organizational imprinting? Cultural entrepreneurship in the founding of the Paris Opera', *American Journal of Sociology*, vol. 113 (1), pp. 97–127.
Katz, M.L. and C. Shapiro (1985) 'Network externalities, competition, and compatibility', *The American Economic Review*, vol. 75, pp. 424–440.
Mahoney, J. (2000) 'Path dependence in historical sociology', *Theory and Society*, vol. 29, pp. 507–548.
March, J.G. (1991) 'Exploration and exploitation in organizational learning', *Organization Science*, vol. 2, pp. 71–87.
March, J.G. (2006) 'Rationality, foolishness, and adaptive intelligence', *Strategic Management Journal*, vol. 27 (3), pp. 201–214.
Martin, R. and P. Sunley (2006) 'Path dependence and regional economic evolution', *Journal of Economic Geography*, vol. 6, pp. 395–437.
Miller, D. (1993) 'The architecture of simplicity', *Academy of Management Review*, vol. 18, pp. 116–138.
North, D.C. (1990) *Institutions, Institutional Change and Economic Performance*. Cambridge: Cambridge University Press.
Ortmann, G. (1995) *Formen der Produktion. Organisation und Rekursivität*. Opladen: Westdeutscher Verlag.
Pierson, P. (2000) 'Increasing returns, path dependence, and the study of politics', *American Political Science Review*, vol. 94, pp. 251–267.
Streeck, W. and K. Thelen (eds) (2005) *Beyond Continuity*. Oxford: Oxford University Press.
Sydow, J., G. Schreyögg and J. Koch (2009) 'Organizational path dependence: Opening the black box', *Academy of Management Review,* vol. 34 (4), pp. 689–709.
Teece, D.J., G. Pisano and A. Shuen (1997) 'Dynamic capabilities and strategic management', *Strategic Management Journal*, vol. 18, pp. 509–533.
Thelen, K. (1999) 'Historical institutionalism and comparative politics', *Annual Review of Political Science*, vol. 2, pp. 369–404.
Witt, U. (1997) '"Lock-ins" vs. "critical masses" – Industrial change under network externalities', *International Journal of Industrial Organization,* vol. 15 (6), pp. 753–773.

2

Slippery Paths of (Mis)Understanding? Historically Based Explanations in Social Science*

Dan Breznitz

2.1 Introduction

This chapter explores why and *how* – path dependence explanations offers us a useful way to analyze the role of history in social change. However for path dependence, to help us to achieve better understanding we must clearly specify it. For these reasons the emphasis in this chapter lies in differentiating and defining the many approaches all which are now confusingly called path dependence, even when each is based on a different set of assumptions about how the world works, and hence, embody different ontology and give us different ways in which to understand social change and conditions for change (Hall, 2003). In so doing this chapter is taking the challenge issued by Paul Pierson, Kathleen Thelen, John L. Campbell, and Colin Crouch to encourage such a debate (Campbell, 2004; Crouch, 2005; Pierson, 2000a; Pierson, 2000b; Thelen, 2004).

The basis for path dependence explanations lies in the assumption that history matters. However, that assumption is usually not defined in any detail beyond the argument that if a choice Y1 is made at time T1, it will affect in some way the set of choices at time T2. As a simple example, if I come home in the afternoon and decide that I will have my Lapsang Souchong with the last piece of lemon in my fridge, I will affect the evening tea choices in the house. We therefore need to make our understanding of path dependence more coherent. In order to so, the chapter is organized as follows. First, I consider the many ways in

which 'path dependence' is used (mayhap overused), and then show how each approach is significantly different in its understanding of social reality. In order to illuminate how each method is rooted in a different ontology, for each of the defined explanatory concepts I present its mechanisms of operation, conditions for social change, and view of social history. I conclude by arguing that we should have a narrow, and hence, much better articulated, definition of path dependence if we are keep the operational utility of the term.

2.1.1 Disentangling path dependence

Most definitions of path dependence are ambiguous, many argue that the choice Y1 taken at the 'critical juncture' time point T1 will lead to a higher probability of taking choice Y2 at time T2 and vastly smaller chances of taking X2 at time T2. In the definition of Paul David, whose work on the development of the QWERTY keyboard set the agenda for the recent revival of the term: 'A path dependence stochastic process is one whose asymptotic distribution evolves as a consequence (function of) the process's own history' (David, 2001).

I contend that there are two main branches of theories in the social science literature that are most commonly named path dependence. One explains how specific outcomes occur as a result of specific sequences of events, it is within these theories that I argue we should search for an accepted definition of path dependence. The other branch of theories, explains how specific institutional systems are resilient over time as a result of the positive feedback they generate, and hence the idea of a specific path, and with it a progression from stage to stage, is less salient, and we should refrain from terming them path dependence. A third, much more disturbing, recent development has been to equate path dependence with any theory that uses past decisions as a static explanation for later outcomes. To elucidate why this is not the case, a brief overview of these approaches and how they differ from path dependence is presented as well.

2.1.2 Path dependence

The original way in which path dependence was used in science was in the natural sciences, describing the specific sequences of things that need to occur to achieve a particular outcome. For example, in order to boil water for my tea I need to (1) fill my pot with water; (2) put it on the stove; (3) light my match; (4) open the gas; (5) light the gas with my match; and (6) wait a few minutes. If I do not perform 6 after 5 after 4 after 3 after 2 after 1, I cannot boil water. Moreover, any change of the

sequence will prevent me from getting the water boiled. Thus, we have three kinds of dependencies – the progression of events is dependent on the path of past events, each event is dependent on the one proceeding it, and the final outcomes are depended on the whole chain (that is path) of events.

A different kind of explanation that social scientists now label as path dependence belongs to the sequencing school. This explanation tries to explain particular outcomes by the specific sequencing of events. For example, if I boil the tea in milk and then add hot water I will get chai; if, on the other end, I first boil the tea in water and then add hot milk I will get tea with milk. A famous example of such theory is Barrington Moore's essay on the social origin of political regimes, where each specific political order of states in the twentieth century – fascism, communism, or democracy – is explained by the sequencing of commercialization and the peasant revolution (Moore, 1966).

Another set of sequencing arguments that are also called path dependence are what Mahoney and others term *reactive sequencing*. These explanations see path dependence as a chain of temporally ordered and causally connected events, where each event in the sequence is both a reaction to the antecedent events and a cause of subsequent events (Mahoney, 2000). Thus, while not exactly following the natural science logic, reactive sequencing explanations advance a very strong claim as to how *a specific path of events,* each leading to the other, leads to *a particular outcome*.

A fascinating, if implicit, use of this view of path dependence as reactive sequence is embedded in John Padget and Christopher Ansell's social network-based research on the rise of the Medici to prominence in fifteenth-century Florence (Padget and Ansell, 1993). Padget and Ansell describe how, through the unplanned creation of a social network with a specific configuration, the Medici were able to have a 'robust action' and be seen by both their 'patricians' and their 'new men' allies as representing their contradicting social ideals and interests. The Medici's social network, it is argued, was created as a reaction to the oligarchs' attempt to punish the Medici for their behavior in a past revolt. First, the oligarchs prevented the Medici from marrying with certain high-status families (a necessity for keeping status and influence in Florence). Second, the oligarchs attempted to strengthen their power in the Medici's own neighborhoods. This caused the Medici to counteract by following a certain pattern of marriage and economic patronism that gave rise to their specific social network. The growth in power of the Medici forced the oligarchs to use more direct methods to try to dispose of the Medici,

who in turn used their social capital by mobilizing their social network, ending the sequence with the Medici ruling Florence.

As such, the sequencing arguments are the closet ones in social science to follow the three different logic of dependencies that first led natural scientists to coin the term: (1) progression of events is dependent on the path of past events – in both sequencing and reactive sequencing, events follow (and are depended) on each other in a specific chain; (2) each event is dependent on the one proceeding it – while this is much clearer in the reactive sequencing, this is true in both cases as a specific sequence of events leads to specific outcome , which of course relate to the third dependence; and (3) the final outcome is depended on the whole chain (that is path) of events.

A third kind of argument now commonly defined as path dependence in social science, is the economic path dependence approach. These explanations argue that path dependence occurs if, after a choice was made for X, the costs of transferring to Y make the move uneconomical, securing the use of X in the future even if Y would be more economically efficient *in the present* if the choices were made in a *tabula-rasa* scenario. In other words, choices at earlier time point T1 leads the system to get 'locked-in,' that is, to enter a specific Nash equilibrium out of other perfectly stable multiple Nash equilibria possible. Here path dependence involves a situation of lock-in after the first choice (David, 1985). Therefore, both the path and dependence allegory are of somewhat limited span in comparison to the sequencing arguments.

2.1.3 Self-reinforcing-sequences

Another set of explanations now termed path dependence are the Self-Reinforcing-Sequences (SRS). These approaches explain the resilience of certain institutional configurations by arguing that the first choice in a sequence creates loops of positive feedback. For example, if Group 1 found that making a specific choice X in decision time point T1 beneficial, Group 1 will tend to keep choosing X for as long as it views X as helping and as long as it possesses the power to do so.

There are three varieties of SRS. Instrumental SRS: these explanations involve an actor or a set of actors benefiting from a specific set of institutions created by a specific choice X in time T1. Those actors will then do everything in their power to keep these beneficial institutions in place in decision time point T2, T3 and so on as long as those institutions are beneficial to them. Thus, if institution A helps Group 1 to gain power, then Group 1 will continue to 'choose' the existence of A.

As a result institution A will stay in existence as long as group 1 is in power and as long as Group 1 perceives institution A as instrumental in its retaining of power.[1] Therefore, instrumental SRS also envision a somewhat shortened path, consisting of only two stages – before and after the construction of the institutions in question. This brings into question the utility of describing them as path dependent.

A second kind of SRS-based explanation is known as critical juncture theory or the punctuated equilibrium approach to social change.[2] This approach claims that major institutional formations, such as the reconstruction of political systems or the reconfiguration of the market, occur in time of crisis. In such periods, the institutional system is more fluid as actors try to find ways to cope with the crisis. A specific institutional system is developed and the crisis passes. Thereafter, the new institutional setting is credited, rightly or wrongly, with this positive outcome, and it takes hold until the next crisis occurs. In addition, as any institutional settings help specific groups to retain or gain power, these groups will then do their utmost to keep these institutions in place, following the instrumental SRS logic. Again, for similar reasons, that is the focus on one critical decision point the utility of terming these explanation path dependent comes into question.

Peter Gourevitch's 1986 book, *Politics in Hard Times,* on the formation of new ruling coalitions under economic crisis is probably the classic exposition of these explanations in political sociology (Gourevitch, 1986). An example of such a research in the field of economic sociology is Neil Fligstein's study of changes in corporate control and behavior in the U.S. between 1919 and 1979 (Fligstein, 1990; Fligstein, 1991). Looking at the 100 biggest U.S. firms at 1919 and 1979, Fligstein finds that in 1919, all these corporations operated only in one market. However, by 1979 about 50 per cent of the biggest American firms were new diversified firms and the rest were old one-market focused firms that had diversified. Fligstein sees three waves of changes, promoted by a new view of what a firm is, in corporate behavior leading to this outcome. Each wave of change was assisted *but not caused* by exogenous crisis, and each crisis was followed by an SRS re-institutionalization of the organizational field.

The third and somewhat less sanguine cause for the persistence of SRSs in history is the Machiavellian approach to social change. Basing his argument on the assumption that change, being a somewhat risky business, frightens people, Machiavelli was the first thinker to urge insightful rulers to present their most extensive reforms as a return to some glorious past. Accordingly, the longer some institutions or

organizations exist, the more they seem 'natural,' and the greater the chances that they will stay (Machiavelli, 1950). The same reasoning is also at the base of another insight, the institutional *bricolage* approach to social change. Building on the original term coined by Claude Levi-Strauss, Giovan Lanzara expanded the meaning of *bricolage* to describe how social institutions are changed not by destruction of older institutions, but by a creative weaving of the old institutions into new ones (Lanzara, 1998; Lévi-Strauss, 1966). Richard Samuels expanded this approach further, arguing that leaders have two basic transformational strategies: 'Bricolage' and 'Revolution.' The *bricoleur* is a leader who actively searches the past in order to shape the present and the future. The *revolutionary* leader rejects a discredited past and tries to explore new forms (Samuels, 2003).

David Stark, in an influential paper that did much to spread the use of the term path dependence in sociology, defines path dependence in a similar fashion (Stark, 1992). Stark argues against the notion advanced by economists in the early 1990s that it is possible to use one 'blueprint' program implemented from above to transform all the Eastern European economies without regard to each country's history. Stark contends that changes, even far-reaching and dramatic, 'are more likely to entail a process of complex reconfiguration of institutional elements rather than their immediate replacement' (Stark, 1992, p. 22). Stark bases his argument on the assumption that when leaders try to depart from long-established routines, they use strategies that resemble these we defined above as Bricolage and Revolutionary. No leader in an existing society, Stark contends, is operating on the optimized clean slate used by economists as the point of departure for their theories.

When analyzing the ontologies of these various theories we find significant differences. Sequencing arguments have both macro and micro views of social history. In their macro view, social history is seen as bringing about the same social dynamics and phenomena, like industrialization, to all societies. However, the temporal ordering of these dynamics influences the final outcome in each society, and actors' potential to change these temporal orderings is seen as limited. In a similar way, the micro view, presented in *reactive sequencing*-inspired work, does not grant agency many possibilities of changing the course of events.

In comparison, SRS explanations are dynamic, granting agency power to bring about social change. Instrumental SRS sees social history as a constant reshuffling of social systems followed by stages of more stable equilibria. Both social stability and social reshuffling are caused by

human agency. This agency is modeled as embedded-rationality guided actors trying to promote and maintain institutions that advance their goals and worldviews and to change institutions that do not. There are two mechanisms for change in the instrumental SRS view. First, when specific benefits that certain institutions grant specific groups diminish, these groups opt to act and change these institutions. Second, new and different groups rise to prominence and opt for institutions that would help them to maintain their dominance and worldviews.

The instrumental SRS view of history is not that different from the punctuated equilibrium view. The main difference being that punctuated equilibrium theory does not grant actors and groups any ability to change the system in its stable phase. Only when an exogenous crisis occurs and weakens the present institutional constellation can actors hope to change the system. Machiavellian institutional view also shares this view of social history as a dynamic process. In this view, groups trying to maximize and maintain their benefits are not the whole story. Leaders creatively trying to shape society and invent new social sense-making dynamics are the main agency of change. However, in a similar way to the instrumental SRS description of history, social change and social innovation have strong dynamic relationships with the past, with the bricolage and the revolutionary leaders using the past as a benchmark and motivator for change.

2.1.4 History matters revisited

Lately, scholars have also started to term many other kinds of historically-based explanations as path dependence. Unlike the two main branches of path dependence which describe present outcomes as the final stage of a specific chain of events, the only common theme of such explanations is the fact that they treat past decisions or non-decisions as given, and explain how their influence is transmitted via the different set of choices and capabilities they bestow on decision-makers at later times.

The most basic approach that argues that history matters in understanding present outcomes contends that past choices affect the material and immaterial resources, which states, organizations, and individuals possess in the present. For example, the fact that IBM decided that in its new venture into microcomputers it would use an operating system owned by Microsoft and a CPU developed by Intel, had significant influence on the future of the three companies and limited the ability of IBM when it decided to write OS2, its own operating system for the PC.

One of the most common ways in which this logic is employed is by arguing that that different institutions give their users different advantages and drawbacks. For example, the German system grants German manufacturing firms high-skilled well-paid labor with rigid hiring and firing laws, while the American system grants American manufacturers very flexible labor laws and low-skilled low-paid manufacturing workers. Thus, when an external crisis hits, different states and organizations might perceive its effects, and consequently the needed response to it, in very different ways. This institution-based approach is now one of the main approaches in comparative political economy under the rubric *types of capitalism*. One example, out of many, is Peter Hall and David Soskice's edited volume, *Varieties of Capitalism: The Institutional Foundations of Comparative Advantage* (Hall and Soskice, 2001). In this volume, the authors claim that nations' different responses to the intensified internationalization of economic activities in the last two decades stem from their different national institutional underpinnings. This volume has created a backlash against the overuse of path-dependence arguments. Indeed the debate around the issue starkly shows the risk of using path-dependence to encompass all varieties of historical explanations of institutional change (Campbell, 2004; Crouch, 2005; Thelen, 2004).

A different strain of this argument contends that past institutional settings influence the present not only because they offer different economic benefits to different strategies, but also because institutions are notoriously hard to change. These writers argue that some institutional resilience can be explained by the fact that institutional systems are vastly complex and intertwined and should be viewed as a whole. Accordingly, even when states or firms try to copy or build similar institutions, the outcomes are vastly different. This line of argumentation is used by many of the writers in Suzanne Berger and Ronald Dore's edited volume, *National Diversity and Global Capitalism* (Berger and Dore, 1996). In the book, the effects of the intensified pressure for convergences of economic activities around the industrial world are shown to be weaker than imagined due to institutional diversity. Wade Jacoby, in his book *Imitation and Politics,* also follows a version of this argument; analyzing the success and failure of Germany's attempt to imitate different industrial relation institutions after WWII and after the reunification of Germany, he shows how in each case the final outcome of an attempt to fully imitate different institutional systems was shaped by the preceding institutions and organizations and the indigenous institutional innovation that evolved around the copied system (Jacoby, 2001).

A different way in which past decisions influence the behavior of states and organizations in the present is the *institutional founding pattern*. This is my, probably too elaborate, name for the fact that formal institutions develops and legitimates a web of informal institutions and 'rules of the game' around them. This web influences the future behavior not only of those organizations and institutions, but also the way in which new formal institutions are constructed in the future. An example for such a research in sociology is Mauro Guillen's book *Models of Management* where he explores differing historical patterns in the adoption of the three major models of organizational management in four countries (Guillen, 1994). Guillen concludes that the difference in adaptation of these models is best explained by the legitimacy of the ideas behind the models and the interplays of these with the other ideologies in society throughout the twentieth century.

While these arguments are similar in being institutionally based, there are important differences in the way they describe social history, consequently, in the mechanisms for change they advance. The varieties of capitalism approach is based on a view of social history in which every society has its own complete set of socio-political-economic institutions that develop slowly through time. As these different institutions grant different economic and industrial advantages and disadvantages, actors usually opt to utilize the specific advantages their particular system grants. Hence, their actions tend to recreate and maintain the existing system.

The view of institutional systems as a gestalt is based on a view of social history in which each and every society creates a thick, heavily linked, and interdependent, institutional system (Jacoby, 2001; Streeck, 1996). As a result, and due to the fact that these systems operate as a whole, even when actors decide to build new institutions or copy existing institutions from different systems, the end result is either very little change, or new institutions intermeshing with the existing system to create a hybrid.

The institutional founding pattern approach is based on a view of social history evolving not only through the creation, maintenance, and destruction of formal institutions, but also through the changing web of informal institutions that surround them and influence their behavior. Changes occur when new formal institutions are created to deal with a specific policy domain. However, the webs of informal institutions and rules of the game that developed around existing formal institutions influence both the ways in which new institutions are created and the future behavior of the old.

In addition to these institutional historically based explanations, there is one more set of widely used explanation that is worth mentioning – *timing*. The timing approach's main argument is that, as history and technology are constantly changing, the timing of actions is important. That is, the point in time when a state or an organization decides to embark on a certain course of action matters. The literature in strategic management about the advantage of first/second/late movers is based firmly in timing-based arguments.

Timing also plays a crucial role in political/socio economy. The whole 'late development' school of thought, growing from Alexander Gerschenkron's seminal work on the role of the state in the economic development of backward countries, is a case in point (Gerschenkron, 1962). A current example of timing explanations is Alice Amsden's contention in her 2001 book *The Rise of The Rest* that the historical moment in which each nation begins its industrialization process determines the kind of industrial revolution path it follows (Amsden, 2001). Amsden conceptualizes industrialization as a three-stage process, with England industrializing on the basis of inventiveness, Western Europe and North America on the basis of innovation, and the rest starting their industrialization after WWII on the basis of learning (Amsden, 2001, p. 2).

Another version of the timing explanation is what Doug McAdam and William Sewell term Master Template in their essay *It's About Time: Temporality in the Study of Social Movement and Revolutions* (McAdam and Sewell, 2001). McAdam and Sewell define 'Master Template' as invented discrete behavioral recipes for making claims, which, once invented, are used by social groups in different locations and periods. The important point to note here is that before the invention of the master template certain claims were not possible, while after the master template is invented they have become a common theme of public discourse. The two examples used by McAdam and Sewell are the French Revolution and the American Civil Movement. These events, once completed, created a recipe for certain political behaviors that have been used by other movements ever since. In a similar way to the historical institutionalism arguments, while timing explanation can be shaped to become path dependence arguments, they inherently have no claim for a specific evolutionary path. Instead their explanatory logic stresses the importance for later development of the particular time in which actions are first taking place.

Table 2.1 presents a summary of all approaches, detailing their mechanisms of operation, conditions for change, and ontology.

Table 2.1 The mixed bag of path dependence, sequencing, and SRS explanations

Approach	**Mechanism**	**Conditions for change**	**Ontology**
Path Dependence			
Natural Science	In order to have a specific output you must follow, in exactly order, certain temporal path of activities.	None.	Laws of nature.
Sequencing	Different sequencing of similar social phenomena and activities produce different and specific outcomes.	Different temporal ordering of similar activities.	Social history brings about major social dynamics/ phenomena that appear in all societies. The temporal ordering of their appearances influences particular societies' structure and welfare. Agency and individual ability to change this temporal ordering are limited.
Economic path dependence	The attaining of stable Nash equilibrium state, through social means.	Certain decisions or activities lead to certain Nash equilibria. Once achieved, these Nash equilibria are very difficult to modify.	Social history seen as a never-ending chain of Nash equilibria. Some of them can be influenced/ decided by social actors, some happen because of historical accidents. Once achieved, most are very difficult to change.

Continued

Table 2.1 Continued

Approach	Mechanism	Conditions for change	Ontology
Self-Reinforcing-Sequences			
Instrumental SRS	Specific institutions benefit specific actors/groups in various ways; these repeatedly maintain/recreate these institutions.	1. The benefits that specific institutions give to certain groups/ actors diminish and these groups opt to change the institutions. 2. Different actors/groups rise to prominence and opt for institutions that would maintain their dominance/ worldview.	Social history is the constant reshuffling of social structures followed by stages of more stable equilibriums. All are caused by embedded-rationality guided actors and groups trying to promote and maintain institutions that advance their goals/worldviews and tear down or change institutions that do not.
Punctuated equilibrium	Exogenous crises hit stable systems and make them fluid for short period of time.	Exogenous crises.	Social history consists of two very different periods: (1) long periods of systems that are inherently stable; and, (2) short periods with major changes that are relatively rare and are caused by exogenous crisis. Actors and groups that hope to implement major changes can do that only under these times of crisis.

Continued

Table 2.1 Continued

Approach	Mechanism	Conditions for change	Ontology
Machiavellian institutional view	Time and tradition lends stronger and stronger legitimacy to institutions transforming them into an inherent part of the stable and common-sense-world and into integral part of the social identity.	Leaders and institutional/ social entrepreneurs force a change on systems, or affect changes after crisis delegitimizes existing institutions. Two modes of operation are bricolage or revolution.	Throughout social history change and social innovation have strong dynamic relationship with the past. Both bricolage and revolution dynamics used the past as benchmark and motivator for both radical and incremental change. History is seen as a thick network of historical links and chains of decisions.
The historical effects of past decisions	Decision to use/ create/develop resources at time point T1 determine the amount of resources available at time point T2	Not of interest	Every decision made in earlier times influences the options available to leaders, firms, and society in later times.
Historical Institutionalism			
Types of capitalism	Different institutions grant different economic and industrial advantages and disadvantages to leaders/ entrepreneurs/ societies when they face similar decision-points. These advantages and disadvantages	Very slow and incremental if at all. Once you have a certain socio-economic institutional system, leaders are inclined to make choices that utilize the specific economic/ industrial advantages their particular system grants. Hence,	Each society has its own complete set of socio-political-economic institutions that change slowly through time. Change is slow and incremental, if at all; what was, is very similar to what is, and also very similar to what will be.

Continued

Table 2.1 Continued

Approach	Mechanism	Conditions for change	Ontology
	make specific decisions/strategies more likely under each particular institutional system.	their actions tend to recreate and maintain the existing system.	Even under crisis the fact that these different institutional systems grant different advantages and disadvantages influence the choices that can be taken by independent agencies/ leaders.
Institutional systems as an intertwined whole	Institutions do not operate in a vacuum but as part of a complete net of intertwined institutions that mitigate and change the behavior and influence of new institutions or of changes to existing institutions.	Decisions to build new institutions or copy existing institutions from different systems usually end with very little change or with the new institutions intermeshing with the existing system to create a hybrid system.	Throughout history each and every society created a thick, heavily linked, and interdependent, institutional system. These systems change only slowly through time by the linking and meshing of new institutions into the existing schema.
Institutional founding pattern	Existence and continuous work of formal intuitions create webs of informal institutions and rules of the game around them. These influence in turn the future behavior of the same institutions and the ways	Formal institutions are created to deal with new socio-political-economic domains and/or implement new policies.	Social history evolved not only through the creation, maintenance and destruction of formal institutions, but also through the changing web of informal institutions that surround them

Continued

Table 2.1 Continued

Approach	Mechanism	Conditions for change	Ontology
	in which new institutions are created.		and influence both their behavior and the formal institutional creation pattern as a whole.
Timing	The particular point of time in which decisions are made and institutions are created makes certain options possible and limits others.	The specific historical context of institutional creation had vast influence on these institutions.	History and technology constantly change. Accordingly, leaders and societies facing the same set of problems at different times have widely different sets of choices open to them.
Master Template	Certain actions create and institutionalize certain social claims and practices. These are then transformed into new templates for social action to be reused and reinterpreted.	From time to time new ideas about social action are institutionalized in such a way that they become a new mode of behavior and/or ordering of the social system that did not exist before. Thus institutionalized, these ideas are diffused and change societies and group behaviors around the world.	Individuals, groups, and organizations are constantly trying to shape and redefine the ways in which society is organized. Their successes and failures as well as the ways in which they do so endlessly change the social structure.
Long Historical Process	The diffusion of ideas and new social forms slowly accumulate to what might seem	Change is constant, if hidden.	The world is persistently changing. This process of changes is continuous

Continued

Table 2.1 Continued

Approach	Mechanism	Conditions for change	Ontology
	as sudden and vast transformations. Many of the seemingly fast changes brought about through crisis are only the culmination of slow processes of social change.		because of new technologies, new ideas, demographic changes, and the tireless attempts of individuals and groups to change their fate and environment.

2.2 Conclusion – is everything path dependent?

In the last few years, the definition of path dependence has been broadened to encompass many different approaches, some of which are based on very different views on how history unfolds. I argue that the differences between all these approaches matter for several reasons. First, while many writers prefer to see all approaches as variations of path dependence, this leads to definitions that are of little value as they do not associate specific casual mechanism and view of social history with path dependence giving the term a floating meaning. Second, as the use of path dependence grows in popularity its definition is constantly being broadened and we run the risk that path dependence will lose all specific causal meaning as a scientific term, and hence, become an empty heading.

The first of these overextended definitions of path dependence associates path dependence with high costs of change, that is, path dependence decisions are decisions that are irreversible or require some cost to change. There are two main drawbacks to this definition. First, if one takes irreversibility seriously, it excludes most social occurrences, as almost no social decisions are completely irreversible. On the other hand if one concentrates on costs of change, then path dependence can be almost every phenomenon that social scientists might have an interest in, since choices that have no long-lasting effects are usually of low interest.

Furthermore, concentrating on cost for change might work very well for economists who do not have dynamic and historic worldview. Indeed,

Paul David argues for defining path dependence as high costs of change, precisely because he thinks that economics has become such an ahistorical and teleological discipline that any approach that might convince economists to account for the effects of history in their explanations is a great improvement (David, 2001). However, for organization scientists, who firmly believe in organization learning and the accumulation of capabilities, focusing on the cost of change is far less useful.

The arena of economic regulations is a clear case in point. If we believe that organizations cannot learn, we might explain regulatory activity as George Stigler did by the inevitable capture of bureaucracies by special interest groups (Stigler, 1971). However, if we do believe that it is the inherent human nature to learn and, therefore, that organization and groups can learn and change as well, we might be more inclined to agree with James Q. Wilson's argument that the changes in the behavior of bureaucratic agencies in the U.S. vis-á-vis special interest groups since 1973 proves that 'Any generalization about how government works is vulnerable to the behavior of persons who have learned that generalization and wish to repeal it' (Wilson, 1980).

Another commonly used definition argues that path dependence is any explanation that advances self-reinforcing-sequences argument. This definition, again, makes the term 'path dependence' redundant. Furthermore, this definition completely disregards the fact that some sequencing arguments are dynamic, that is, this definition sees social agencies as having an ability to bring about change. However, some are static, so once a critical decision has been made in the far mists of time, social agents have no way of changing it.

I contend that a better approach is to utilize a much more precise definition of path dependence. For that end, I argue that *sequencing* and *reactive sequencing* is the best way to view true path dependence. Sequencing is the claim that each outcome is dependent on the specific set (that is sequence) of events proceeding it. Reactive sequence is the claim that specific outcomes are caused by a chain of temporally ordered and causally connected events, where each event in the sequence is both a reaction to the antecedent events and a cause of subsequent events. This definition refocuses our attention on the fact that other explanations pertain to the resilience of specific institutions and how they change throughout time, and does not aim to explain how taking a specific path leads to specific outcome. Furthermore, of all the explanations now termed path dependence by social scientists – whether sequencing or reactive sequencing – present an explanation that it inherently builds on the notion of a path of events. It relates only

to the three different dependencies of these paths: dependence in progression, dependencies between events, and dependencies of the final outcome on the specific path of events.

Indeed, one rich venue for future research on path dependence is to follow exactly this line of thought and to inquire into the ways in which the use of different explanations influences analysis and conclusions about the same social phenomena, even when there is an agreement on the historical facts. Hence, while this chapter limits itself to the notion that there is real need to define, clarify, and define path dependence in order to improve the quality of historically based social science research, social scientists may only hope that future researchers will not stop here, but expand our understanding of how the uses of different approaches inherently affects our understanding of the social phenomena.

Notes

* This paper was first written as part of the work with the History Collaborative Research Network of the SSRC program on the Corporation as a Social Institution, and was supported through grant given by the Sloan Foundation. The author would like to thank the participants of the History CRN workshop in MIT – Center for International Studies in April 2002, especially, Doug Gutherie, Daniel Friel, Ajay Mehrotra, Katherine Chen, Rohit Daniel Wadhwani, and Liba Brent; the participants in Harvard's Center for European Studies Comparative Historical Analysis Lunch series of 2001–2002, especially Hillel Soifer, Fiona Barker, and Shannon O'Neil; Suzanne Berger and Richard Samuels for reading and rereading of this paper and for encouraging him to further develop it in every step of the way; to Peter Perdue, Amos Zehavi, Michal Ben-Josef Hirsch, Brett V. Kubicek, Elizabeth Popp Berman, Cory Welt, and Stephen Van Evera for insightful analysis and useful remarks. Special thanks are also owed to Roy Moxham's book Tea: Addiction, Explorations, and Empire. All mistakes are solely the author's.

1. James Mahoney argues that there are four logics, each leading to a different kind of what I termed here instrumental SRS: (1) Utilitarian Explanations: groups or organizations keep choosing X because it benefits them; (2) Functional: institutions are reproduced because they serve a certain function for the larger system in which they are embedded; (3) Power: similar to utilitarian explanations but arguing that institutions distribute benefits unevenly so the resilience of certain institutions might be intimately linked to the supremacy of certain elites; and (4) Legitimation Explanations: institutions are being reproduced because they fit a certain ideology or view of how the world should be shaped, a perspective that is held by a significant number of members of a society (Mahoney, 2000).
2. Stephen D. Krasner coined the term 'punctuated equilibrium' in his 1984 article, honoring the theory in evolutionary biology that brought him to these insights (Krasner, 1984).

References

Amsden, A. (2001) *The Rise of 'The Rest' Challenges to the West from Late Industrializing Economies*. Oxford: Oxford University Press.

Berger, S. and R. Dore (1996) *National Diversity and Global Capitalism*. Ithaca: Cornell University Press.

Campbell, J.L. (2004) *Institutional Change and Globalization*. Princeton: Princeton University Press.

Crouch, C. (2005) *Capitalist Diversity and Change: Recombinant Governance and Institutional Entrepreneurs*. Oxford: Oxford University Press.

David, P.A. (1985) 'Clio and the economics of QWERTY', *American Economic Review*, vol. 75, pp. 332–337.

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

Fligstein, N. (1990) *The Transformation of Corporate Control*. Cambridge, MA: Harvard University Press.

Fligstein, N. (1991) 'The structural transformation of American industry: An institutional account of the causes of diversification in the largest firms, 1919–1979', in W.W. Powell and P. DiMaggio (eds) *The New Institutionalism in Organizational Analysis*. Chicago: The University of Chicago Press, pp. 311–336.

Gerschenkron, A. (1962) *Economic Backwardness in Historical Perspective, a Book of Essays*. Cambridge, MA: Belknap Press of Harvard University Press.

Gourevitch, P. (1986) *Politics in Hard Times*. Ithaca: Cornell University Press.

Guillen, F.M. (1994) *Models of Management: Work, Authority, and Organization in a Comparative Perspective*. Chicago: University of Chicago Press.

Hall, P. (2003) 'Aligning ontology and methodology in comparative research', in J. Mahoney and D. Rueschemeyer (eds) *Comparative Historical Analysis in the Social Sciences*. Cambridge, U.K.: Cambridge University Press, pp. 373–406.

Hall, P. and D. Soskice (2001) *Varieties of Capitalism: The Institutional Foundations of Comparative Advantage*. Oxford: Oxford University Press.

Jacoby, W. (2001) *Imitation and Politics*. Ithaca: Cornell University Press.

Krasner, S.D. (1984) 'Approaches to the state: Alternative conceptions and historical dynamics', *Comparative Politics*, vol. 16, pp. 223–246.

Lanzara, G.F. (1998) 'Self-destructive processes in institution building and some modest countervailing mechanisms', *European Journal of Political Research*, vol. 33, pp. 1–39.

Lévi-Strauss, C. (1966) *The Savage Mind*. London: Weidenfeld & Nicolson.

Machiavelli, N. (1950) *The Prince and the Discourses*. New York: Random House.

Mahoney, J. (2000) 'Path dependence in historical sociology', *Theory and Society*, vol. 29, pp. 507–549.

McAdam, D. and W.H.J. Sewell (2001) 'It's about time: Temporality in the study of social movement and revolutions', in R. Aminzade, J. Goldstone, D. McAdam, E.J. Perry, W.H.J. Sewell, S. Tarrow and C. Tilly (eds) *Silence and Voice in the Study of Contentious Politics*. Cambridge U.K.: Cambridge University Press, pp. 89–125.

Moore, B.J. (1966) *Social Origins of Dictatorships and Democracy: Lord and Peasant in the Making of the Modern World*. Boston: Beacon Press.

Padget, F.J. and K.C. Ansell (1993) 'Robust action and the rise of the Medici, 1400–1434', *American Journal of Sociology*, vol. 98, pp. 1259–1319.
Pierson, P. (2000a) 'Increasing returns, path dependence, and the study of politics', *APSR*, vol. 94, pp. 251–268.
Pierson, P. (2000b) 'Not just what but when: Timing and sequence in political processes', *Studies in American Political Development*, Spring, pp. 72–92.
Samuels, R.J. (2003) *Machiavelli's Children: Leaders and Their Legacies in Italy and Japan*. Ithaca: Cornell University Press.
Stark, D. (1992) 'Path dependence and privatization strategies in east central Europe', *East European Politics and Societies*, vol. 6, pp. 17–54.
Stigler, G.J. (1971) 'The theory of economic regulation', *Bell Journal of Economics and Management Science*, vol. 2, pp. 3–21.
Streeck, W. (1996) 'Lean production in the German automobile industry: A test case for convergence theory', in S. Berger and R. Dore (eds) *National Diversity and Global Capitalism*. Ithaca: Cornell University Press, pp. 138–170.
Thelen, K. (2004) *How Institutions Evolve: The Political Economy of Skills in Germany, Britain, the United States and Japan*. New York: Cambridge University Press.
Wilson, J.Q. (1980) 'The politics of regulation', in J.Q. Wilson (ed.) *The Politics of Regulation*. New York: Basic Books, pp. 82–103.

Part II

Path Dependence in Public and Private Organizations

3

Imprinting, Path Dependence and Metaroutines: The Genesis and Development of the Toyota Production System*

Hugo van Driel and Wilfred Dolfsma

3.1 Introduction

The notion of path dependence was first explicitly applied in the field of economics, in particular to explain the persistence of certain technologies and standards (David, 1985; Arthur, 1989). It has been discussed in recent decades in other disciplines as well, and has been found to be a useful way of analyzing a range of subjects. The concept has become particularly popular in sociology and political science, often aiming to explain institutional development (Pierson, 2004; Mahoney, 2000). At the micro level of individual organizations much less use of the concept has been made so far, with the notable exception of David (1994). Only recently has the interest in this field of application expanded (e.g., Sydow et al., 2009). Applying the idea of path dependence to the analysis of organizational change raises major theoretical issues, however. We illustrate this with a re-analysis of the origin and development of the Toyota Production System.

Analyzing this case from the perspective of path dependence leads us to a number of observations of both an empirical as well as a theoretical nature. First and foremost, initial conditions that are of vital importance in path-dependent processes are ill-defined in the literature. We show that assessing both the content and timing of initial conditions is a major research task. For Toyota, the most relevant initial conditions are not the often-cited external conditions relating to demand and resources, but rather its specific *metaroutine* of 'self-testing and adapting'

developed over a long period of time by Sakichi Toyoda, active as an inventor and entrepreneur in the textile machinery industry between 1895 and 1930, before Toyota was actually founded by his son Kiichiro in 1937. The metaroutine has committed Toyota to its typical production system despite changing conditions. Toyota's main Japanese rival Nissan, facing the same circumstances, chose a different route from the 1930s onwards.

As a consequence, secondly, we underline the importance of 'imprinting' by initial conditions. This chapter is not a plea for a deterministic, one-dimensional understanding of path-dependence processes, however. In the case of Toyota, a metaroutine made certain choices more salient than other ones, but did not entail them. Imprinting by this metaroutine was a multifaceted process, noticeable at the initial adoption of the system, but also during its further evolution and it occurred partly through affiliated operational routines. In addition, a particular sequence of events during the critical years of 1949 and 1950 induced Toyota to diffuse its then still infant production system within the company and finally also to suppliers. A focus on metaroutines illuminates how imprinting 'resonates' with specific events, creating dynamic lock-in. We thus steer a middle road between overly deterministic and overly voluntaristic interpretations of organizational change.

3.2 Initial conditions and imprinting in organizational path dependence

In its common interpretation, whether using these exact terms or not, path dependence consists of a combination of sensitivity to initial conditions and reinforcing mechanisms creating lock-in. The latter are usually explicitly defined in the path-dependence literature. Beyer (2005), for instance, distinguishes several mechanisms creating continuity: increasing returns, sequences of events, functionality, complementarity, power, legitimacy and conformity. The path-dependence literature is less elaborate about the nature of initial conditions. Noteworthy is a recent comment on the related concept of 'critical juncture' as applied in political science: 'The paucity of conceptual instruments available to define, study, and compare critical junctures is striking when compared with the rich conceptual apparatus (e.g., increasing returns, lock-in, sequencing) used to analyze path-dependent processes themselves' (Capoccia and Kelemen, 2007, p. 343).

We aim to address this shortcoming. In particular, the selection mechanisms as part of the initial conditions need to be more clearly

conceptualized. The customary assumption in the path-dependence literature is that the direction a path takes under initial conditions is largely undetermined (e.g., Arthur, 1989; Dopfer, 1991; Goldstone, 1998; Mahoney, 2000). This is contested by Sydow et al. who point out that the stage before a path starts is not characterized by 'completely unrestricted choice' (Sydow et al., 2009, p. 693). We build on this and understand (early) 'imprinting' (Stinchcombe, 1965; Marquis, 2003) explicitly as part of a path-dependence process. Slightly reframing a categorization by Kriauciunas and Kale (2006), the initial conditions potentially having an imprinting effect can be grouped into three categories: (1) environmental requirements; (2) resources; and (3) values or philosophies. Of these categories, 'philosophies' deserve some further elaboration. Philosophies may conceptually be conceived of as an abstract type of routine. In studying path dependence, higher order routines are of special interest, since they both modify existing routines and guide the search for and the selection of new routines (Nelson and Winter, 1982, p. 18). They thus open up the possibility of explaining dynamics and change. We refer to these higher-order routines as 'metaroutines.' Routines do not need to be formalized or recognized, explicitly or implicitly in organizations (Nelson and Winter 1982), and so we would like to avoid exclusive connotations of the term 'metaroutines' with formal procedures for standardized problem solving as has been done by Adler et al. (1999), or with very specific phenomena such as production lines (Szulanski and Jensen, 2004).

Following a suggestion by Becker (2004, p. 664) to link behavioral and cognitive interpretations of routines, we primarily understand metaroutines as a propensity to select particular solutions for certain types of problems. Industry- or economy-wide routines can be held responsible for the time-sensitive social 'imprinting' (Stinchcombe, 1965; Marquis, 2003).[1] Alternatively, firm-specific metaroutines – similar to organization-specific 'codes' (David, 1994) – may leave their imprint. As they are not so easily imitated by others, imprinting will often occur at the mundane level only, and thus contribute to the variety in choices of firms that are confronted with more or less similar external conditions.

Imprinting is not pre-determined, since it always involves interpretation and conscious action by agents (Johnson, 2007; Djelic and Quack, 2007). Analogous to an observation made about political processes as being different from economic ones in regard to path dependence (Capoccia and Kelemen, 2007, pp. 353–354), acts and decisions by individual agents may be considered a category of 'events' of special importance in directing organizational paths (see also Garud and Karnøe, 2001).

Consciously acting upon initial conditions, this agency makes imprinting by certain directions of action and choice more likely than others.[2] However, imprinting is unlikely to be a 'one-shot' phenomenon, since initial conditions will disappear only unevenly over time, gradually or suddenly. At the same time, this partial change implies that repeated imprinting is not simply a full-scale replication. Moreover, organizational features themselves, as studied here, are – unlike, for instance, (basic) technical standards – multi-dimensional phenomena that are not created or selected in their entirety at a certain distinct point of time (see Thelen, 2004 and Capoccia and Kelemen, 2007, pp. 349–350, regarding institutional change).

The process of repeated imprinting thus implies several possible links between initial conditions and a lock-in, where a lock-in is a dynamic phenomenon, representing continuity and not rigidity. Taking this perspective enables us to include in our analysis in a natural way the role of (unforeseen) sequences of events that figure so prominently in many path-dependence interpretations. They can both act as a contingent factor for initial selection and as a reinforcing mechanism later on, in varying degrees as distinct from (repeated) imprinting by initial conditions. These theoretical considerations give rise to several questions for empirical research. Which relevant initial conditions can be specified in the form of environmental requirements, resources and metaroutines? How and to what extent do they (repeatedly) imprint the phenomenon under study? How do these imprinting forces interact with (sequences of) events in selecting and locking-in certain choices and action patterns in an organization?

The origin and development of the often-studied Toyota Production System (TPS) seems an appropriate case for such an empirical application. The TPS, which originates from the mid-twentieth century, is one of the most famous production systems. Its two closely related main elements are (1) 'lean production,' of which small lot-size production using a just-in-time (JIT) system is the most well-known feature, and (2) the use of a both flexible and integrated multi-layered network of suppliers. After seeing the success of the system (Womack et al., 1990), other car makers, both in- and outside Japan, have tried to emulate Toyota's approach, but have certainly not completely succeeded in doing so. Toyota has been considered the best-organized and most productive carmaker in the world for decades. Leading longitudinal studies of the TPS (Cusumano 1985; Fujimoto, 1999; Coriat, 2000) take an implicit path-dependence perspective and so cannot fully exploit the concept as developed here. Supplementing this historical case study,

we compare Toyota with its main domestic rival Nissan, a comparison that is often made in the literature. Such a comparison can serve the same purpose as the more problematic, yet regularly proposed method of counterfactual history as part of a path-dependence analysis (Booth, 2003; MacKay, 2007).

Given limited space, we cannot discuss all elements of the TPS mentioned above and will restrict our focus to internal production. Our empirical sources consist of the extensive secondary literature on Toyota and Nissan published in English. For the sake of convenience, we group our analysis under two headings: demand (the key environmental requirement in this case) and resources, on the one hand, and metaroutines, on the other. Sensitivity to initial conditions and reinforcing mechanisms are respectively dealt with in these sections, including the role of specific (sequences of) events.

3.3 Path dependence in the genesis and development of TPS

3.3.1 Demand and resources: continuity in change

The prevailing view in leading studies on the Japanese car industry is that both a low, fragmented demand and scarce resources have called for a lean production system (Cusumano, 1985; Ohno, 1988; Fujimoto, 1999; Coriat, 2000). The literature thus presents two major potential initial conditions of the TPS. However, in precisely dating the start of the TPS, and thus demarcating initial conditions from the temporal context under which reinforcing mechanisms are supposed to function, this does not appear to be self-evident. One year after founding Toyota, Kiichiro Toyoda (1894–1952), commenced JIT production in Toyota's brand new Koromo factory in September 1938 (Wada and Yui, 2002). However, the uninterrupted implementation and diffusion of the TPS originates from around 1948, when engineer Taiichi Ohno (1912–1990), who is widely considered the father of the TPS, began his experiments with 'lean production' in Toyota's machine shop (Ohno, 1988; Cusumano, 1985, pp. 262–307). Fortunately, an alternative dating of initial conditions does not make a large difference to our analysis of their actual effects that leads us to largely reject the customary view.

Both in 1938 and around 1948, domestic Japanese car makers faced a demand for vehicles that was low, but not truly fragmented. In 1938 it consisted mainly of military trucks (Cusumano, 1985; Wada and Yui, 2002); Japan had entered into a full-scale war with China the year before. Around 1948, when the U.S. was effectively ruling the country through

its General Headquarters (GHC; 1945–1952), domestic production of private cars was still virtually forbidden and the Japanese car makers almost exclusively served a limited market for fairly standardized civilian trucks (Cusumano, 1985, p. 266; Toyota, 1988, p. 461). Regarding resources, Toyota's very limited financial means have affected its production system as a capital saving one. More empathically, however, scarcity of material resources has been quoted as a typical Japanese circumstance inviting for lean production. This type of scarcity indeed already plagued Japan in the war economy of 1938 and even more so around 1948. However, shortage of materials, leading to unreliability in their supply, hampered rather than stimulated the use of JIT at Toyota from 1939 onwards (Cusumano, 1985, p. 278; Toyota, 1988, p. 142).

Ironically, Ohno focused his efforts to reduce 'waste' on labor (Ohno, 1988; Shingo, 1989), the most widely available resource, as many workers lost their jobs after Japan surrendered to the allied forces in August 1945 (Gordon, 1985, pp. 334 and 346). Nevertheless, Ohno tried to reduce Toyota's dependence on skilled labor in particular. This is understandable, since the workers had improved their negotiation position in one important respect in the first post-war years. Stimulated by the GHQ, membership of labor unions exploded from virtually non-existing to 6.5 million in June 1948 (Okayama, 1987, p. 171) and a nationwide autoworkers' union was founded in 1947. Ohno was thus responding to changing labor relations rather than to constraints in resources or to the specific nature of demand for automobiles in Japan.

Both the nature of demand and the availability of resources changed from 1948 in a comparatively short period of time due to a sequence of events, not completely transforming initial conditions, but intensifying the need to align production to demand and economize on costs as Toyota's already weak financial position deteriorated dramatically. The government first stimulated the Japanese car makers to expand production in the second half of 1948, but subsequently curbed demand through its anti-inflation measures in April 1949. As a result, the stock of unsold vehicles (still mainly trucks) mounted, and Toyota faced a grave liquidity crisis by the end of 1949. Banks saved the company from bankruptcy, but forced Toyota to split up into a production company (Toyota Motor) and a sales company (Toyota Sales) in April 1950 (Toyota, 1988, pp. 92, 104–105). This separation, discontinued only in 1982, was meant to prevent overproduction in the future. The 'pull' principle that a car was to be produced only when the sales company ordered one 'resonated' with the JIT principle pioneered by Ohno (Togo and Wartman, 1993, p. 103). Subsequently, the banks also

pressed Toyota to lay off one third of its labor force, which it grudgingly did in June 1950, when approximately 2,000 'volunteers' left the company. Simultaneously, an exogenous shock, a sudden boom in U.S. demand for military trucks after the start of the Korean War in June 1950, boosted the Japanese car makers' production. Toyota realized this increase in production with a significantly reduced labor force, calling for an extension of labor-saving measures, such as multi-machine handling by one operative, a major element of lean production introduced by Ohno in the machine shop.

A contingent combination of unexpected changes in demand thus eventually contributed to Toyota's lock-in into its typical production system, initially adopted under different demand conditions. But this start of diffusion of lean production within Toyota antedated the emergence of a still low-volume, but increasingly diversified Japanese market for private cars in the course of the 1950s (Cusumano, 1985, p. 266) that is normally seen as the condition underlying the TPS. Likewise, despite its development into a mass producer of cars in the 1960s, Toyota only selectively and temporarily went over to the typical U.S. practice of long production runs of single components or cars (Cusumano, 1985, pp. 280–281; Toyota, 1988, pp. 326–327). By then, if it had wanted to invest heavily in dedicated machinery and large intermediate stocks, Toyota also would no longer have been hampered by limited financial resources (Cusumano, 1985, p. 76).

3.3.2 The dominating influence of metaroutines

Although facing the same circumstances, Nissan took a different route to developing its production system. Next to the crucial impact of the intervention of the banks at Toyota in 1949–1950, this hints at the importance of metaroutines in shaping the TPS. The typical Toyota metaroutine of 'self-testing and adapting' originated from Sakichi Toyoda (1867–1930). As the son of a carpenter, he became obsessed with improving the hand looms he saw his mother and other women use in the textile cottage industry in the Aichi region. Sakichi's concerns throughout his career were securing enough funds to finance his inventions and improvements of textile machines, and creating opportunities to test them in practice. He founded, owned, restructured, and/or managed a range of firms in the textile and textile machinery business between 1895 and his death in 1930. The poor performance of his first power loom in comparison with British Platt Bross looms in a one-year-lasting mill experiment by the Kanegafuchi Cotton Spinning Co around 1907 (Wada and Yui, 2002, pp. 26–27) taught Sakichi to develop

and test his innovations under (his own) mill conditions (Mass and Robertson, 1996, p. 6; Toyota, 1988, p. 30).

This was the start of the 'self-testing and adapting' metaroutine of Toyota. In 1930, several years before the founding of the Toyota Motor Company (1937), Sakichi's son Kiichiro – who had graduated as an engineer – started experimenting with manufacturing gasoline engines under the roof of Toyoda Automatic Loom Works, the main textile machinery venture created by his father. That the father would 'imprint' his son's mentality was not inevitable. Initially, Sakichi had a career in mind for his son as an 'entrepreneur' not to be involved in lengthy factory experiments (Wada and Yui, 2002). From his side, as a teenager Kiichiro is said to have been so irritated by all the public praise for his father's work that he insisted towards his good friends at school that 'Anyone who worked by trial and error ... could not be a great inventor' (Togo and Wartman, 1991, p. 30). Nevertheless, according to Kiichiro himself, because of a critical experience with testing 200 looms together with his father in 1924 (note the similarity with the latter's own experience in 1907), he eventually took over Sakichi's typical philosophy of 'actually trying comes first' (Wada and Yui, 2002, pp. 129–130). The adoption of Toyota's lean production system should be seen against the background of this metaroutine. Kiichiro's early attempt in 1938 to use a JIT system at the Koromo plant primarily as a device to identify bottlenecks in production (Wada and Yui, 2002, pp. 288–289) was in line with this general focus on unabated testing and adapting. Similarly, it fitted well with Ohno's permanent obsession with removing 'waste.'

The contrast with Nissan illustrates that the Toyota approach was not self-evident under the prevailing circumstances. Through a complicated story of mergers and restructurings, Nissan became involved in the production of Datsun cars in 1933. Nissan was the abbreviation that came into use to denote the Nippon Sangyo holding firms created by Yoshisuke Aikawa in 1928 (on Nissan, see Cusumano, 1985, and Odaka et al., 1988). Aikawa admired U.S. big business and had spent most of his time between 1905 and 1909 in the U.S. Although he, an engineer like Kiichiro, deliberately started his working life on the shop floor (Odaka et al., 1988, p. 93), Aikawa is described as someone 'who was more interested in building empires than cars' (Cusumano, 1985, p. 52). Nissan produced Datsun cars and parts for the then dominant U.S. car plants in Japan under the guidance of American engineers from 1935, which was in fact a continuation of the cooperation between Aikawa with William R. Gorham and other U.S. engineers dating back to 1920

(Odaka et al., 1988, pp. 95–96). When after a few years Nissan shifted its focus to truck production in response to demand from the military, Aikawa bought an entire truck factory in the United States, moved it to Japan and began operations based upon the accompanying designs and with the help of American engineers (Cusumano, 1985, p. 27). In contrast to this 'quick copy' approach, Kiichiro and his collaborators – supported by university people – tried to master the production of parts themselves by various lengthy experiments. Between September 1933, when a separate 'automotive department' of Toyoda Automatic Loom was established, and May 1935 not a single car was actually produced (Odaka et al., 1988, pp. 125–130; Wada and Yui, 2002, pp. 235–254). This indicates both the care Kiichiro took to master production processes himself and the large problems involved with it.

Toyota's metaroutine outlasted the initial conditions and thus was vital in sustaining the company's production system. The repeated 'imprinting' by the metaroutine differed in specificity, concerned different aspects and occurred through both 'old' and 'new' mechanisms, so it went beyond pure replication. More generally, not only Kiichiro, but also his cousin Eiji was stimulated by Sakichi to keep in touch with work-floor practices and insights during and after his engineering studies (Toyota, 1988, p. 38; Wada and Yui, 2002, p.130). After Kiichiro had resigned from Toyota in 1950, Eiji became the main protector of Ohno, who encountered widespread internal resistance against his lean production approach (Ohno, 1988, p. 36). Sakichi's focus on self-testing and adapting machines also influenced later developments at Toyota more specifically through the culmination of his efforts in the famous type-G loom of 1924 – the first automated loom with a non-stop automated shuttle change. Multi-machine operating (see above) had been common already at Toyoda Weaving and Spinning, where Ohno started his career in 1933, before he applied it at Toyota from around 1948. This practice required that machines automatically stop in cases of trouble or at the end of the production process; a typical 'self-testing' feature for which the G-loom acted as a template.

Similarly, Ohno later adapted older multi-purpose equipment to 'form repetitive operations almost automatically [...] on the principle of the Toyoda automatic loom' (Cusumano, 1985, p. 275). Furthermore, obviously, Kiichiro's initial adoption of a JIT approach inspired (and legitimized) Ohno's post-war experiments (Ohno, 1988). Also crucial was Kiichiro's habit of writing down in very detailed reports findings of his self-testing and adapting experiments, and manuals for production and organization (Wada and Yui, 2002). Such formalization by

the president himself was highly unusual among Japanese companies (Hino, 2006, p. 10). In combination with Toyota's policy of transferring knowledge through the circulation of personnel within plants, with Ohno as the most prominent example, this extensive recording of experience was vital for the enduring effect of the metaroutine of 'self-testing and adapting.' More specifically, the rule that workers can stop the lines when JIT production becomes in danger or when other problems occur, which was first introduced in the engine shop in 1950 (Cusumano, 1985, p. 280), can also be considered an inbuilt-testing mechanism helping to sustain the application of Toyota's metaroutine of self-testing and adapting.

To be sure, like Sakichi, both Kiichiro and Ohno held American production methods and equipment in high esteem.[3] However, they never simply copied or bought them as package. Instead they used them as 'unbundled technologies' (Fujimoto, 1999, p. 38), adapting them for their own purposes. Typically, Toyota adopted American-made specialized tools much more selectively than Nissan from the 1950s onwards (Okayama, 1987, p. 178; Daito, 2000). One might speculate whether the timing of specific events contributed to locking-in Toyota to its production system in this respect. The outbreak of the Korean War led the U.S. government to frustrate international technical cooperation of the kind Toyota and Ford were about to agree on in June 1950 (Kamiya, 1976, pp. 71–72). In 1952–1953, when the company was no longer in crisis, Toyota, unlike the majority of the Japanese car makers, did not enter into a 'tie-up' with a Western automaker, despite an encouragement to do so from the Japanese government (Cusumano, 1985, p. 8). Faithful to its prevailing 'quick copy' approach, Nissan tried to build cars exactly to the blueprints of the production of its British partner Austin (Cusumano, 1985, p. 100). The tie-ups did not significantly contribute to the other car makers taking a different path than Toyota; however, – as in line with the Japanese government's policy – complete 'domestication' of production by Nissan, Isuzu and Hino was reached by 1958 (Odaka et al., 1988, p. 45).

Like Toyota, Nissan began to apply JIT-like features from the mid-1950s (Cusumano, 1985, pp. 307–319). Nevertheless, differences in the pervasiveness and persistence in implementation remained important and Toyota consistently outperformed Nissan in sales, productivity and also often in quality from the 1950s onwards (Cusumano, 1985, p. 392, pp. 396–397; Odaka et al., 1988, pp. 91, 108–109; Shimokawa, 1987, p. 229). As late as the mid-1980s, Toyota's efforts in continuously improving operations were still more oriented to the organization of processes compared to Nissan's stronger focus on machines

and equipment (Totsuka, 1995; Fujimoto, 1999, p. 69). An important example of a more recently observed mechanism apparently reflecting Toyota's metaroutine is that a 'core group of shop-floor leaders function as an informal intra-organizational mechanism for pre-screening the routines' (Fujimoto, 1999, p. 275). This allows Toyota to 'quickly convert a variety of organizational elements into a coherent system' (Fujimoto, 1999, p. 264). To conclude, currently, in a much more refined and advanced form than at the time of founding, Toyota's metaroutine of 'self-testing and adapting' still prevails. Toyota's general approach is thus typified in recent studies by 'built-in tests to signal problems automatically' (Spear and Bowen, 1999, p. 98) or 'integration of problem identification and problem-solving procedures into the actual work processes' (Towill, 2007, p. 3620).

3.4 Conclusion

Applying the concept of path dependence to the Toyota Production System (TPS) forces scholars to specify the nature of the historical explanation of organizational change. Firstly, we argue for the necessity of opening the black box of initial conditions as part of a path-dependence analysis. A careful assessment of the exact timing and plausible effect of the (supposed) initial conditions helps to understand whether and how they shaped the phenomenon under study. Scarcity of material resources and a limited, fragmented domestic demand for cars (environmental requirements) were much less decisive initial conditions than normally suggested in the literature on the genesis of the TPS, also considering the different path taken by Nissan starting from the same situation. On the other hand, a shortage of financial means, but, above all, firm-specific metaroutines were of particular relevance in imprinting organizational features during initial conditions. A metaroutine of 'self-testing and adapting,' that was developed by the father of Toyota's founder and taken over by the latter, in both cases particularly after having a critical experience with testing equipment, made the just-in-time approach an obvious, although not necessary choice. This 'imprinting' counters the often supposed indeterminate relation between initial conditions and the selection of the path taken.

Secondly, the metaroutine repeatedly imprinted onto the organization; Toyota's philosophy of 'self-testing and adapting' continued to shape the TPS, so that it developed as a uniquely identifiable production system. This was not a matter of simple replication, however, as both initial conditions and the phenomenon studied itself evolved as constituting elements changing gradually or suddenly. Re-imprinting

thus happened in a changing context, concerned several aspects of the product system and for an important part took place through affiliated – pre-existing and newly created – operational routines. There were also continuities between the other initial conditions and the circumstances prevailing during the critical years of 1949–1950. Still, the culmination of the sequence of events in a crucial intervention by the banks acted as a clearly distinct reinforcing mechanism in inducing Toyota into an internal diffusion of its emerging production system that was first adopted under partly different conditions of demand and resources. Studying the TPS thus illustrates how metaroutines, acting as imprinting mechanisms, can shape organizations both during initial conditions as well as in subsequent periods, in a path-dependent interplay with certain other initial conditions, evolving operational routines, and crucial sequences of events. The lock-in was dynamic in nature, making the TPS a continuously changing production system that has been hard to emulate by competitors.

3.5 Epilogue

Recent developments at Toyota underline the need for steering a middle road between overly deterministic and overly voluntaristic interpretations of change of organizations. A more 'aggressive' policy by Toyota's top management has become observable since 1995, when Hiroshi Okuda became CEO. This policy change was partly contested by members of the Toyoda family (Hino, 2006). It shows that Toyota's metaroutine is not immune from erosion: philosophies at the top are starting to change, weakening existing metaroutines and possibly replacing them with others. This change in mood seems to have led Toyota to deal with outside suppliers in a more arm's length manner (Ahmadjian and Lincoln, 2001). For its production methods as such, despite adjustments, the general principles of the TPS are still upheld (Benders and Morita, 2004; Nohara, 1999; Pil and Fujimoto, 2007).

The current CEO, Katsuaki Watanabe recently begun to speak about a 'kakushin' (revolution) needed at Toyota, as opposed to the traditional 'kaizen' approach of incremental improvement (Stewart and Raman, 2007). There is reason for such a change, since, at the point of taking over the position of the largest car producer in the world from ailing General Motors, Toyota has become confronted with serious quality problems in production (Anonymous, 2007). Therefore, in 2006, Watanabe ordered that time pressure by the sales department on 'product development' and 'manufacturing' to turn out new car models should be lessened,

so that enough time would become available to prepare the production of defect-free cars. This might suggest a break from Toyota's sales-led policy originating from the critical 1949–1950 period (Treece, 2007). Whether or not the TPS will be affected at its heart as a result of these (indicated) recent policy changes remains to be seen.

Notes

***** This chapter builds on our 'Path dependence, Initial conditions, and Routines in Organizations: the Toyota Production System Re-examined,' *Journal of Organizational Change Management*, 2009, vol. 22 (1), pp. 49–72.

1. Unlike Kriauciunas and Kale (2006), we thus consider not only resources, but also values or philosophies as potentially both external and internal to the firm.
2. Note that, conversely, the discretion of agents under initial conditions might lead to an (overly) voluntaristic interpretation of sensitivity to initial conditions (Capoccia and Kelemen, 2007, p. 347).
3. In addition, the 'just-in-time' re-stocking of empty American supermarket shelves served as an important source of inspiration for Ohno (Ohno, 1988).

References

Adler, P.S., B. Goldoftas and D.I. Levine (1999) 'Flexibility versus efficiency? A case study of model changeover in the Toyota production system', *Organization Science*, vol. 10, pp. 43–68.

Ahmadjian, C.L. and J.R. Lincoln (2001) 'Keiretsu, governance, and learning: Case studies in change for the Japanese automotive industry', *Organization Science*, vol. 12, pp. 683–701.

Anonymous (2007) 'Briefing Toyota. A wobble on the road to the top', *The Economist*, November 10th–16th, pp. 85–87.

Arthur, B.A. (1989) 'Competing technologies, increasing returns, and lock-in by historical events', *Economic Journal*, vol. 99, pp. 116–131.

Becker, M.C. (2004) 'Organizational routines', *Industrial and Corporate Change*, vol. 13, pp. 643–677.

Benders, J. and M. Morita (2004) 'Changes in Toyota motors' operations management', *International Journal of Production Research*, vol. 3, pp. 433–444.

Beyer, J. (2005) 'Pfadabhängigkeit ist nicht gleich Pfadabhängigkeit. Wider den impliziten Konservatismus eines gängigen konzepts', *Zeitschrift für Soziologie*, vol. 34, pp. 5–21.

Booth, C. (2003) 'Does history matter in strategy? The possibilities and problems of counterfactual analysis', *Management Decision*, vol. 41, pp. 96–104.

Capoccia, G. and R.D. Kelemen (2007) 'The study of critical junctures. Theory, narrative, and counterfactuals in historical institutionalism', *World Politics*, vol. 59, pp. 341–369.

Coriat, B. (2000) 'The "abominable Ohno production system." Competences, monitoring, and routines in Japanese production systems', in G. Dosi,

R.R. Nelson and S.G. Winter (eds) *The Nature and Dynamics of Organizational Capabilities*. Oxford: Oxford University Press, pp. 213–243.
Cusumano, M.A. (1985) *The Japanese Automobile Industry*. Cambridge, MA: Harvard University Press.
Daito, E. (2000) 'Automation and the organization of production in the Japanese automobile industry', *Enterprise & Society*, vol. 1, pp. 139–178.
David, P.A. (1985) 'Clio and the economics of QWERTY', *American Economic Review*, vol. 75, pp. 332–337.
David, P.A. (1994) 'Why are institutions the "carriers of History"? Path dependence and the evolution of conventions, organizations and institutions', *Structural Change and Economic Dynamics*, vol. 5, pp. 205–220.
Djelic, M.L. and S. Quack (2007) 'Overcoming path dependency: Path generation in open systems', *Theory and Society*, vol. 36, pp. 161–186.
Dopfer, K. (1991) 'Toward a theory of economic institutions: Synergy and path dependency', *Journal of Economic Issues*, vol. 25, pp. 535–550.
Fujimoto, T. (1999) *The Evolution of a Manufacturing System at Toyota*. Oxford: Oxford University Press.
Garud, R. and P. Karnøe (2001) 'Path creation as a process of mindful deviation', in R. Garud and P. Karnøe (eds) *Path Dependence and Creation*. Mahway: Erlbaum, pp. 1–36.
Goldstone, J. (1998) 'Initial conditions, general laws, path-dependence, and explanation in historical sociology', *American Journal of Sociology*, vol. 104, pp. 829–845.
Gordon, A. (1985) *The Evolution of Labor Relations in Japan : Heavy Industry, 1853–1955*. Cambridge, MA: Harvard University Press.
Hino, S. (2006) *Inside the Mind of Toyota*, translation of Japanese edition from 2002. New York: Productivity Press.
Johnson, V. (2007) 'What is organizational imprinting? Cultural entrepreneurship in the founding of the Paris Opera', *American Journal of Sociology*, vol. 113, pp. 97–117.
Kamiya, S. (1976) *My Life with Toyota*, translation. Toyota Motor Sales Company.
Kriauciunas, A. and P. Kale (2006) 'The impact of socialist imprinting and search on resource change: A study of firms in lithuania', *Strategic Management Journal*, vol. 27, pp. 659–679.
Mackay, R.B. (2007) '"What if?": Synthesizing debates and advancing prospects of using virtual history in management and organization theory', *Management & Organizational History*, vol. 2, pp. 295–314.
Mahoney, J. (2000) 'Path dependence in historical sociology', *Theory and Society*, vol. 29, pp. 507–548.
Marquis, C. (2003) 'The pressure of the past: Network imprinting in intercorporate communities', *Administrative Science Quarterly*, vol. 48, pp. 655–689.
Mass, W. and A. Robertson (1996) 'From textiles to automobiles: Mechanical and organizational innovation in the Toyoda enterprises, 1895–1933', *Business and Economic History*, vol. 25 (2), pp. 1–37.
Nelson, R.R. and S.G. Winter (1982) *An Evolutionary Theory of Economic Change*. Cambridge, MA: The Belknap Press.
Nohar, H. (1999) 'The historic reversal of the division of labour? The second stage of the Toyota production system', in J.P. Durand, P. Stewart, J.J. Castillo (eds) *Teamwork in the Automobile Industry*. Houndmills: Macmillan, pp. 37–53.

Odaka, K., K. Ono and F. Adachi (1988) *The Automobile Industry in Japan*. Tokyo: Kinokuniya Company Ltd.

Ohno, T. (1988) *Toyota Production System*, translation from Japanese edition of 1978. Cambridge, MA: Productivity Press.

Okayama, R. (1987) 'Industrial relations in the Japanese automobile industry 1945–1970: The case of Toyota', in S. Tolliday and J. Zeitlin (eds) *The Automobile Industry and Its Workers*. New York: St. Martin's Press, pp. 168–189.

Pierson, P. (2004) *Politics in Time*. Princeton and Oxford: Princeton University Press.

Pil, F.K. and T. Fujimoto (2007) 'Lean and reflective production: The dynamic nature of production models', *International Journal of Production Research*, vol. 45, pp. 3741–3761.

Shimokawa, K. (1987) 'Product and labour strategies in Japan', in S. Tolliday and J. Zeitlin (eds) *The Automobile Industry and Its Workers*. New York: St. Martin's Press, pp. 224–243.

Shingo, S. (1989) *A Study of the Toyota Production System from an Industrial Engineering Viewpoint*. Cambridge, MA: Productivity Press.

Spear, S. and H.K. Bowen (1999) 'Decoding the DNA of the Toyota production system', *Harvard Business Review*, vol. 77 (5), pp. 97–106.

Stewart, T.A. and A.P. Raman (2007) 'Lessons from Toyota's long drive', *Harvard Business Review*, vol. 85 (7/8), pp. 74–83.

Stinchcombe, A.L. (1965) 'Social structure and organizations', in J.G. March (ed.) *Handbook of Organizations*. Chicago: Rand McNally, pp. 142–193.

Sydow, J., G. Schreyögg and J. Koch (2009) 'Organizational paths: Path dependence: Opening the black box', *Academy of Management Review,* vol. 34 (4), pp. 689–709.

Szulanski, G. and R.J. Jensen (2004) 'Overcoming stickiness: An empirical investigation of the role of the template in the replication of organizational routines', *Managerial and Decision Economics*, vol. 25, pp. 347–363.

Thelen, K. (2004) *How Institutions Evolve: The Political Economy of Skills in Germany, Britain, the United States, and Japan*. Cambridge: Cambridge University Press.

Togo, Y. and W. Wartman (1993) *Against All Odds. The Story of the Toyota Motor Corporation and the Family that Created It*. New York: St. Martin's Press.

Totsuka, H. (1995) 'The transformation of Japanese industrial relations: A case study of the automobile industry', in S. Babson (ed.) *Lean Work. Empowerment and Exploitation in the Global Auto Industry*. Detroit: Wayne State University Press, pp. 108–128.

Toyota. A History of the First 50 Years (1988) translation. Toyota City: Toyota Motor Corporation.

Towill D.R. (2007) 'Exploiting the DNA of the Toyota production system', *International Journal of Production Research*, vol. 45, pp. 3619–3637.

Treece, J.B. (2007) 'Quality problems lead to soul searching: Are we growing too fast?', *Automotive News*, vol. 82, Supplement October 10, pp. 178–179.

Wada, K. and T. Yui (2002) *Courage and Change. The Life of Kiichiro Toyoda*, translated. Toyota City: Toyota Motor Corporation.

Womack, J.P., D.T. Jones and D. Jones (1990) *The Machine that Changed the World*. New York: MacMillan.

4

Path Dependence, Institutionalization and the Decline of Two Public Institutions*

Sanneke Kuipers and Arjen Boin

4.1 The price of success?

The early decades of the last century witnessed the rise of two fascinating public institutions in the U.S. At the federal level, the Tennessee Valley Authority came to prominence. On the east coast, the Port of New York Authority – dormant for years following its formal inception – blossomed into an immensely powerful organization. Intriguingly, both organizations also experienced a painful decline. Both organizations still exist, but their heydays seem long past.

Some organizations manage to build on early successes and stay at the top of their game for decades (Collins and Porras, 1994). Others become the victim of their own success: they stick to previously successful routines and practices, eventually succumbing to the detrimental effects of conservative inclinations. These organizations somehow lose their 'institutional fit' with their environment.

Path dependence theory may shed light on the decline of successful institutions (cf. Carter, 2008). Path dependence theory holds that the mechanisms that bring performance stability and external support, 'lock-in' further organizational development. It suggests that the roots of institutionalization can put an organization on a trajectory from which deviation becomes increasingly difficult over time. Path dependence may thus account for the organization's lack of adaptive capacity. The very path-dependent mechanisms that help an organization institutionalize in its early years may bring about its institutional demise years down the road.

In this chapter, we explore if and how the initial steps of young public organizations determine future trajectories of institutional development (Mahoney, 2000; Pierson, 2004; cf. Kuipers, 2006). We investigate whether the sources of institutional decline stem from the organization's early years, and ask if and how path dependence can lead to critical phases of rapid deinstitutionalization. To answer this question we revisit the institutional histories of the Tennessee Valley Authority (TVA) and the New York Port Authority. Both organizations symbolized 'creative efforts to fashion new instruments of government to meet economic and social needs' that cross state boundaries and have a specialized regional function (Doig, 2001, p. 1). Both organizations suffered institutional demise in later life. Yet, we see also striking differences: the powerful organizational design and the early institutionalization of the TVA (Hargrove, 1994; Colignon, 1997) provide a sharp contrast with the vague and toothless charter, and the slow institutionalization, of the Port Authority (Doig, 2001; Walsh, 1975).

We begin by discussing theoretical approaches to path dependence and institutionalization, and explore the possible relation between lock-in processes and institutional collapse. Sections 4.3 and 4.4 analyze the role of path-dependent mechanisms in bringing about the institutional demise of the Tennessee Valley Authority and the Port Authority of New York and New Jersey. Our approach is informed by the process-tracing method, which draws our attention to a select set of variables and maps their development in time (George, 1979). We conclude with an agenda for further research.

4.2 The downside of institutionalization

When young organizations design or discover ways of operating that prove successful, they typically build their organization around what has become their 'core technology' (cf. Thompson, 1967). By structuring the organization around such a core technology, it may get caught in a 'competency trap' (Levitt and March, 1988): what works today may not work tomorrow (as organizational environments change). Previous successes can easily blind an organization to emerging gaps between their competencies and the evolving environmental demands (Jervis, 1997).

Blinded by success, such organizations inadvertently yet vigorously pursue the road towards deinstitutionalization (Miles, 1980).[1] Success – here defined in terms of high levels of legitimacy and goal attainment – not only saps the adaptive capacity of an organization, it may give rise to hubris. Few tasks are then seen as beyond reach; retrenchment is felt

to be an abdication of responsibility. Highly institutionalized organizations seem particularly vulnerable to hubris. Their distinct identity and the values they embody, make institutions special in the eyes of their members and their audiences. This makes it hard to deviate from traditional and often cherished practices and paradigms.

A core idea underlying the path-dependence perspective is that each initial investment increases the costs of switching to an alternative way of working (Pierson, 2000). Over time, adaptation to environmental changes becomes more difficult. If we add to this equation the capacity for self-deception, it becomes easy to believe that successful organizations are doomed to fail (Landau and Chisholm, 1995).

An organization becomes crisis-prone when rigidity has been building up during long stretches of path-dependent organizational development. It is therefore imperative to study path dependence over extensive periods of time. The path-dependence phenomenon is caused by lock-in mechanisms or so-called institutional reproduction.

Three mechanisms of institutional reproduction appear especially relevant (see for a detailed discussion of all four mechanisms: Mahoney, 2000).[2] First, the *power resources mechanism* enhances the power of organizational elites through their exclusive access to key resources. Such resources are information and expertise, privileged access to decision-making venues, and a means to accumulate organizational assets such as capital or members (i.e. mobilization pressure on other decision makers). Actors with dissenting views often do not have such resources or influence. As a result, dissenting opinions, criticism and unfavorable change of circumstance may go unnoticed in the organization for a long time. In addition, those in power are unlikely to change the organizational conditions from which they benefit.

Second, the *efficiency mechanism* refers to comparative advantages of the existing structure against forthcoming alternatives, which Pierson (2000, p. 252) calls 'increasing returns.' Increasing returns occur 'when the relative benefits of an activity compared with other options increase over time.' Sunk costs and acquired rights will make abolishment of the existing system – in which money, expertise, and time have been invested – very expensive. The proposed alternative not only requires additional investments, but also has uncertain benefits. Many stakeholders favor the shortcomings of the status quo above undefined future advantages.

Third, the *legitimacy mechanism* explains how an organization becomes consolidated in its environment. This happens when the organization's key actors come to dominate the definition and understanding of the organization's tasks. Protagonists of the organization thus establish the

standards against which the organization's performance will be measured. The organization can ignore outside complaints and criticism as long as it controls the evaluation of its performance among its most important stakeholders. This attitude becomes problematic when the organization overestimates how its self-defined standards are valued, or when it underestimates outside criticism.

Institutional reproduction furthers the natural process of institutional erosion, which, if unchecked, creates a gap between the institution's practices and the inevitably changing environment (cf. Hay, 2001). Organizational environments always change, as new issues become politically important, perceptions alter and levels of tolerance are renegotiated. The gap between a self-absorbed organization and its environment can grow, and later present the organization's decision makers with problems they can no longer solve by the usual means. This gap provides fertile ground for institutional decline, and – ultimately – the occurrence of an institutional crisis (Boin and 't Hart, 2001; cf. Alink et al., 2001; Kuipers, 2004).

For each of the mechanisms, we have constructed a hypothesis that will guide the case analysis. We will use these hypotheses to investigate why the early success of the two organizations under study eventually gave rise to institutional decline.

HYPOTHESIS 1: *The more resources and influence the elites of a (sub)organization have acquired, the more they are able to expand their influence over the organization's future course, even if this drives the organization away from its original tasks and external legitimacy.*

HYPOTHESIS 2: *The more money, expertise and time that have been invested in certain goals and outputs of the organization, the more these investments will come to dominate strategic policy choices of the organization, even in the face of evidence that additional investments will not lead to positive returns.*

HYPOTHESIS 3: *The more legitimacy an organization enjoys, the more it defines the standards by which the judgment of external actors takes place, which lays the organization open to criticism when it can no longer meet its own evaluation criteria due to changing circumstances.*

4.3 The Tennessee Valley Authority

The Tennessee Valley Authority (TVA), established in 1933, was 'a corporation clothed with the power of government but possessed of the

flexibility and initiative of a private enterprise' (Hargrove, 1994, p. 20, quoting Franklin D. Roosevelt).[3] Its mission was to improve the economic situation of the under-developed Tennessee Valley, a region encompassing seven states. TVA had to protect the region against flooding and soil erosion while at the same time exploit hydroelectric power. TVA built sixteen dams and hydroelectric facilities between 1933 and 1944, it protected the Valley against floods of the Tennessee River, generated hydroelectricity, set a national yardstick for electricity rates, employed tens of thousands of workers, and taught the people of the Valley how to fertilize and conserve their land.

Fifty years later the organization seemed to combine the worst features of business and government: while democratic accountability and public legitimacy were virtually absent, the organization was unconstrained in its spending and in the exercise of its authority. In 1985, TVA had to close down five nuclear reactors because they did not meet the safety requirements of the Nuclear Regulatory Commission.[4] Other reactors were either cancelled, deferred or under construction: none were operating. The nuclear program had cost $14 billion and the people of the Valley paid $1 million extra per day for their utilities, without receiving extra electricity. To make things worse, TVA had exposed the Valley to grave nuclear hazards due to unsafe operations. As a result, the TVA became the subject of intense public and Congressional scrutiny and criticism. Two consecutive chairmen were demoted, 40 per cent of the staff was laid off and the TVA suffered a drastic reorganization. What had happened to the crown jewel of the New Deal? The section below presents an analysis of path dependence by looking at the effect of three reproductive mechanisms in the TVA case.

4.3.1 Power: from instrument to end in itself

The power resources mechanism explains both the TVA's increasing competitive advantage in the private market and the TVA's increasing resistance to political control. New demands and criticisms by stakeholders in later years fell on deaf ears in an organization that had become so independent of sponsors. Internally, the organization became increasingly divided. The TVA department responsible for generating and selling electricity, the Power Office, dominated the others and could neither be controlled within the organization nor from the outside.

The TVA was created in 1934 as a public corporation, which meant that it enjoyed considerable autonomy from federal interference. The tripartite board had the discretion to act, aside from final approval by the Bureau of the Budget, as it wished within the context of the Tennessee

Valley Authority Act (Hargrove, 1994, p. 106). The TVA invested federal funds in building hydroelectric facilities, and employed the revenues from the sale of electricity to further strengthen its position in the Valley vis-à-vis the private utilities.

Even though control of the Tennessee River had been the main aim of the TVA Act (electrical power was a by-product), TVA soon became *the* electric power company for the Tennessee Valley. Congress consented with TVA's autonomy because the organization was self-supporting through the sale of electricity. Vice versa, the autonomy granted by Congress allowed TVA to make investments to strengthen its independent status (Colignon, 1997, p. 275). The passage in 1959 of the Bond Revenue Act further enabled TVA to finance its power program through the sale of bonds (Hargrove, 1994, p. 117).[5] This allowed TVA to finance steam plants, and later, nuclear facilities. Clearly, power had become more than a by-product of river control.

Opposition from state and local politics was limited. The TVA effectively used its status as a cross-state authority to successfully deny state regulation. The Authority enjoyed national recognition for its mission to develop an economically backward region. Over time, however, TVA increasingly became a regional agency with private business-like interests. This shift from national to regional mission was not matched by a shift in mechanisms of accountability (Hargrove, 1994, pp. 299–300). Whereas Congress and the Office of Management and Budget lost their grip on the TVA, democratic institutions at the state level did not gain control: the TVA operated in a political vacuum.

Neither Congress nor the OMB thoroughly scrutinized the TVA's expansion plans to supply electricity generated by steam and nuclear plants. Instead, Congress raised the TVA's debt ceiling to unwarranted levels in the 1970s to cover the expenses of the construction of nuclear facilities. Congressional delegates from the Valley continued for a long time to support the TVA. The delegates first acted when discontent rose in the Valley about the TVA's electricity rate increases. Rates had increased by 580 per cent in 15 years prior to 1981 – partly because of rising costs of energy in general, but also because of the TVA's unsound spending on the nuclear program. The 1985 nuclear debacle would arouse the anger of the rest of Congress.

Internally, the TVA was becoming increasingly divided. At the time of creation, Congress had given little consideration to the TVA's administrative design. The new organization was unique in its formal objective of integrating multiple purposes: there was no pattern, no experience, and no plan (Hargrove, 1994, p. 24). An organizational structure

emerged that matched the TVA's mission of the first two decades: vigorous action to alter the physical environment. The board decided on the broad outlines and delegated the work to skilled technicians to assess the feasibility and desirability of particular plans. Conflict among the board members of the 1930s later drove the different divisions – each operating under the authority of a different board member – apart (Colignon, 1997).

The TVA organization was driven by operational concerns: pragmatic solutions required decentralization of authority, professional discretion, on-site decision-making, and little central budgetary control. This situation was unproblematic as long as there was sufficient money to spend. When these conditions dissipated, the TVA comprised of separate largely autonomous offices, unwilling to give up their decentralized authority in the face of new environmental demands and objections against the nuclear program.

The TVA was not at all up for the task of building nuclear plants. The organizational structure had not changed in 50 years, but the hazardous technology it operated certainly had. Dams and steam plants in the early decades could be built in a hurry, with high levels of discretion among engineers on the site. The complexity of nuclear plant construction required much more time, intense cooperation between offices, and strict compliance with external regulation. The TVA's historically grown inability to meet such requirements made the construction of nuclear plants an inefficient, ineffective and plainly unsafe affair.

4.3.2 The downside of efficiency: elasticity works both ways

The hydroelectric facilities initially benefitted most people in the valley. The TVA provided them with cheap electricity. The electricity prices set by the TVA served as a yardstick to scrutinize private utilities' electricity sales in the rest of the country.

Two practical maxims followed from this early success, which later 'became almost religious tenets' (Hargrove, 1994, p. 7). First, TVA officials believed that the cheaper the cost to the consumer, the greater the use of electricity. In the early years of the power program this certainly was true, for both residential and industrial consumers. Second, TVA policy makers claimed that supply of electricity should always precede the demand for it, in order to keep up continuous economic development of the valley.

The TVA worked hard to stimulate the demand for electricity. It created a public consumer credit affiliate, the Electric Home and Farm

Authority, to provide low-interest loans to farmers for the purchase of appliances (Hargrove, 1994, p. 45). Chairman David Lilienthal stated that 'there is no such thing as too much electrical energy [...] Electrical energy creates its own market' (Hargrove, 1994, p. 125). The TVA justified the increase of electricity supply (for which more power plants had to be built) by spreading the gospel that economic development of the Valley was dependent on the availability of cheap electricity. Power supply should stay ahead of the demand in order to create growth. When from the 1940s onwards the interests of national defense also justified increasing supply, the tail was beginning to wag the dog (Hargrove, 1994, p. 63 and pp. 125–127).

The electricity program began to lead a life of its own. The TVA could easily take the step from building hydroelectric facilities to building steam plants. After all, the TVA's own engineers and construction workers were ready to go to work – and who would want them to lose their jobs? The TVA's board later argued likewise regarding the construction of nuclear plants, even though nuclear technology proved to be much more difficult (Hargrove, 1994, p. 126). Yet when the TVA board realized it had been overambitious, sunken costs and commitments to workers made reversal of earlier decisions politically infeasible (Hargrove, 1994, pp. 244–245).

At first, nuclear energy seemed a great alternative source of power. In response to complaints about air pollution, and in response to the rising prices during the oil crises of the early 1970s, the TVA board felt pressed to find new sources of cheap energy – the basis of TVA success. Board chairman Aubrey Wagner thought that nuclear power would be cheaper, cleaner and less controversial than coal (Hargrove, 1994, p. 127).

Historically, the sale of electricity enabled the TVA to create more plants, supply more electricity, yield more revenues, and again create more power-generating facilities to keep up with the increased demand for electricity. The TVA board considered its ambitious plan to build seventeen nuclear reactors feasible because it assumed that demand for electricity would continue to rise as long as it was preceded by ample supply. However, TVA engineers miscalculated and underestimated the costs of building nuclear plants. With the costs rising, the price elasticity of energy showed its other side: demand for electricity snapped back. Industries, private customers and other government agencies reduced their demand for TVA electricity. Distributors threatened to buy their electricity from other utilities, which would leave the TVA with a huge debt and no customers. Efficiency concerns had thus driven the TVA

further down the path of increasing energy supply, and efficiency concerns later drove its customers away.

4.3.3 When the legitimating myth no longer matched reality

The TVA leadership created an 'agency myth' through which the organization legitimated its actions and existence (Hargrove and Glidewell, 1990). This myth proved 'extremely useful for the socialization of TVA employees, as a defense against outside control, and as a genuine guide to several of the TVA operations' (Hargrove, 1994, p. 6). The myth evolved around the idea of grassroots democracy, which meant that the TVA involved the people of the Valley in setting the aims and ambitions of the Authority. The TVA had to be primarily a service to the people of the Valley, rather than a political tool for Washington DC or for state governments. This idea permeated the TVA's personnel policy, its relations with local stakeholders, and its stance towards Congress and the President.

The effect of the grassroots idea on employees was that bottom-up decision-making became highly valued within the TVA. The inculcation of this grassroots myth would for decades prevent the board from exercising much control over the operating divisions, even when increasingly complex technology and regulation warranted such control (Hargrove, 1994, p. 102). Internally, the grassroots 'democracy' reduced centralized control, with different TVA offices increasingly going their own way.

Externally, the grassroots ideal became the standard TVA defense against encroachment from Washington (Hargrove, 1994, p. 53). The TVA presented itself as an organization for and by the people, and any interference as an attack on the citizens of the Valley. Whether the obstacle was the White House, Congress or the private utilities, the TVA successfully mobilized public support against the opposition (Hargrove, 1994, p. 57).

The absence of political control and the TVA's status as the benchmark for other utility companies, made the organization virtually free of regulation in setting its rates and evaluating its own performance. This freedom would later allow the TVA to increase its own rates drastically to cover the expenses of its nuclear expansion. The TVA had decided to build steam plants and nuclear reactors because the organization felt it had to provide cheap electricity. Cheap electricity, in turn had been the basis for public support.

The two legitimating factors – the idea of grassroots democracy and low-cost electricity – had become the standard by which people judged

the TVA. The Authority lost its grassroots support when the people in the Valley would no longer favor its programs and policies. People no longer favored its programs and policies when the TVA could no longer provide inexpensive electricity. The basis for the TVA's legitimacy thus began to crumble as a consequence of the TVA's efforts to maintain this legitimacy.

4.4 The Port of New York Authority[6]

The Port of New York Authority (PA, currently named Port Authority of New York and New Jersey) is a bi-state agency that was created in 1921 to increase cooperation and economic development on both sides of the Hudson River. Its aim was to aid commercial activity and transport mobility related to one of the world's largest seaports, located on the border between the states New York and New Jersey, and between the cities New York and Newark. The agency was born out of an interstate conflict about the standardized railroad freight rates, which benefited New York over New Jersey. The PA's specific mandate was to solve the conflict and improve the region's rail-freight system. At the same time, the new agency was given a very broad mission: to enhance interstate cooperative planning in general. However, the PA received neither the funding nor the power to carry out either its general mission or its explicit task.

The PA's path of organizational development would be a rugged one. In the 1920s, the Authority essentially failed in its task to implement a comprehensive rail-freight plan, and ten years after its creation the rail transport situation and the unequal economic development on both sides of the Hudson were still as bad as they had been in 1921 (Walsh, 1978, p. 98). The Great Depression drove the Authority close to bankruptcy. The conservative Board decided to rid the Authority of its only successful department (engineering) and direct its focus towards repaying its debts by maintaining and operating existing bridges and tunnels. By 1941, two decades after its inception, the PA 'sat passively on dead center' (Bard, 1942, p. 266).

In the mid-1940s the Port Authority rose from the brink of death, when Austin Tobin became its executive director. Tobin built an 'empire on the Hudson,' expanding the Authority's domain to harbour and airport development, bus and truck terminals, in addition to building bridges and tunnels. During Tobin's reign of almost 30 years, the Authority's staff tripled in size and its revenues became disproportionately larger than the organization could absorb in its existing projects.

The PA therefore decided to build the World Trade Centre twin towers and was forced to invest in a New Jersey rail transit line. The twin towers would soon come to be seen as symbolic of the Authority's increasing arrogance.

Political discontent with Tobin's growing insensitivity to political concerns, community protests and environmentalist demands drove him into sudden early retirement in 1972. After Tobin stepped down, the Authority's activities, mission and staff morale slowly disintegrated under consecutive new directors. The agency entered the twenty-first century riddled with interstate distrust and conflict, patronage and political staff appointments; still in search of a new comprehensive mission. The sections below present a path-dependent analysis of the PA case by looking at the effect institutional reproduction, through power resources, efficiency and legitimacy.

4.4.1 The mixed blessings of divided powers

Political conflict and historical distrust between New York and New Jersey meant that the PA could not act without approval of both state legislations. The governors of both states could always exercise their veto. In addition, the cities, the counties, the judiciary and the federal government had various options to thwart the Authority's plans. This made the Authority's autonomy contingent upon cooperation with many political actors.

When the PA sought to force the railroad companies into cooperation in the 1920s, New Jersey legislators blocked their efforts because they feared that New York would benefit most from the rail improvement. When state governors appointed incapable board members in the 1930s, there was little the Authority could do. Likewise, the states could force the Authority in the 1960s to invest in the ailing railroad system, and deprive the organization of its control over its financial surpluses in the 1990s.

The PA's power was not self-reinforcing but fluctuating, conditioned by politically strategic leadership on the Authority's part, and politically supporting leadership in its environment. With a cunning strategist such as Austin Tobin (1942–1972) at the helm, the PA managed to survive external attacks such as city planner Robert Moses' obstruction of the PA's airport and bus-terminal ambitions. In order to realize its ambitions, the Authority formed ad-hoc coalitions with those actors serving its purpose. For instance, the national airlines had at first actively endorsed the Authority's conquest to take over the responsibility for the region's three main airports from the cities. Once the airports

were in the Authority's 'hands,' Tobin turned against his former allies to break their monopoly in setting the rates for using airport facilities. He forged a nationwide coalition with other airports who suffered from the airlines' 'below cost' rental rates.

Though Tobin managed to bestow internal cohesion and a strong sense of purpose on the organization,[7] he could not free the Authority from its dependence on political actors and financial currents. Tobin could embark on new building projects, which in turn generated revenues for even more projects. Yet, approval of state legislators and governors was required for each new project. Moreover, political actors could impose tasks – such as rail improvement – on the Authority that would drain its financial reserves, as it did from the late 1960s onwards.

4.4.2 The efficiency of an empty vessel

The PA's broad mission – to enhance cooperative planning, transport mobility and economic development in the bi-state region – could include almost any activity. There were virtually no limits on the tasks that political actors could impose upon it. Whenever an interstate issue came up, state politicians could ask the Authority to take it on. Doig (2001, p. 11) describes the Authority as an 'empty vessel, into which various ideas for interstate cooperation on transportation issues might be tossed'. What would be more efficient than using an agency that is already up and running?

Sometimes this political delegation was beneficial to the growth and institutionalization of the PA. For instance, when Newark city officials asked it to take on airport improvement in 1945, the Authority was able to press New York City officials to hand over responsibility for New York's airports as well (Kaufman, 1952). Initiative from the Authority's political environment would not always be a welcome spur to its activities, though. Whenever the railroad problems were cast in its lap, the Authority seemed to run into trouble (Walsh, 1978, pp. 97–98). In the 1920s, the Authority's mandate to improve the rail system became a fiasco. In the 1960s, the Board became responsible for the ailing PATH rail transit system.

The organization, therefore, actively worked its environment to receive the mandates it preferred. Austin Tobin was a master strategist, and guided the Authority successfully. The PA had to 'structure the world so [it] could win' (Riker, 1986, p. ix). Whenever the Authority would stay passive, the organization would be no more than a toll collector. When the Authority became too active in manipulating its environment, it invited criticism (Walsh, 1978, p. 295). This was the

case when the Authority wished to build a fourth airport in 1959. The organization 'failed to comprehend the range and depth of resistance' to its plans, and in public hearings Port Authority officials conveyed an impression of arrogance (Doig, 2001, pp. 385–387). The jetport would be the first project Tobin failed to force through, and his efforts to do so would herald the start of declining legitimacy for his organization.

The PA's expertise and existing infrastructure could lead to new assignments, but not necessarily so. Sunken costs guaranteed that the Authority would always have work in maintaining and operating the existing infrastructures. But the Authority remained heavily dependent on the strategic behavior of its executive officials and opportunities arising from its environment.

4.4.3 The self-supporting principle

One of the early established maxims of the PA was that construction projects had to be financially self-supporting in the long run through bonds and revenues. Financial surpluses generated by the revenues of one project could help finance the early deficits or investment costs of the next project. As such, the PA legitimized its existence by conducting self-financing projects that would benefit the larger bi-state metropolitan area.

However, sometimes the Authority was forced to invest in rail projects that would never be self-financing (Walsh, 1978). The Authority set this self-financing principle as a standard by which it justified its projects; yet its statutes implied that others could define which projects the Authority had to adopt. The ailing PATH rail project was a cuckoo's egg that the state legislators forced the Authority to accept. Yet because the PA had always proudly asserted that its projects were self-supporting, its financial performance was judged on that basis – even if the Authority now had to account for rail-transit deficits that were not its fault. During the 1960s, the Authority gradually lost support for its existing achievements because it had to raise the toll rates to cover the expenses of the ailing railroad system.

In addition, the Authority's obsession with the self-financing and efficiency aspects of existing and new construction projects undermined democratic accountability (Doig, 2001). The PA often clashed with local communities in its efforts to initiate new projects. The Authority allowed financial estimates and efficiency considerations to overshadow the importance of public support, such as when it set out to build the WTC (Walsh, 1978, p. 103). The construction of the WTC exceeded its budget by one billion dollars: 'a monument to the city's former glory as

a port and to the Authority's audacious ability to get its own way' (Doig, 2001, p. 385, quoting the *The New York Times*).

The self-supporting principle was the key to many of the Authority's early successes. The Authority could expand its domain indefinitely as long as projects – be they airports, harbors, bus terminals, or bridges – would generate sufficient revenue to meet their own total costs. However, the more the Authority used this financing principle as a justification for its expansion, the more vulnerable it became to criticism when it could no longer meet its own evaluation criteria in the late 1960s.

4.5 Does institutionalization create path dependence?

This chapter considered if and why the institutionalization of public organizations can eventually cause institutional decline. Public organizations that wish to be successful, typically translate their legislative mandate into a set of critical tasks: tasks that are feasible, that can be effectively fulfilled and that generate achievements that keep the organization's stakeholder content (Wilson, 1989). Path dependence theory predicts that the way an organization builds itself around its critical tasks has a defining impact on its future development. The three hypotheses, derived from path dependence theory, indeed seem to explain the fate of the TVA quite well.

First, the TVA translated its broad mission into the task to supply inexpensive electricity. Providing power was both the key to the organization's early institutionalization, and later to its institutional demise. The TVA invested revenues generated by power sales in more power-generating capacity. When the production of electricity and its consequences became less appreciated, the tail began to wag the dog. The TVA board was unable to change its course in time.

Second, the TVA had long been successful in enlarging the scale of production to lower electricity prices. Unfortunately, economies of scale also drove the TVA board to deciding that generating nuclear power would be the most efficient way to keep supply up and cost down. However, due to underestimation of the complexity and cost of nuclear technology, and due to changing political circumstances, electricity rates went up and demand for electricity went down.

Third, the grassroots myth that made the TVA so successful in legitimizing its existence and performance only worked well as long as interests did not collide. Changes in the organization's environment required centralization of authoritative decision-making to change the organization's course. These new demands triggered an organizational

identity crisis for TVA employees, and a legitimacy crisis vis-à-vis the people of the Valley.

The Port Authority's mission was as equally broad as the TVA's mission: to increase cooperation and economic development on both sides of the Hudson River. Yet the PA also received a specific task: improve the rail-freight system. It was not until the 1940s that Austin Tobin aligned the mission, tasks, resources and external legitimacy of the New York PA into one coherent, valued enterprise. While institutionalization was slow to come about, it went quickly after Tobin took the reign. The seeds of decline were institutionalized as well, however.

The Authority's statutes gave several political actors in both states veto power over its plans. Over time, the PA won credibility and a reputation as an organization that 'got things done.' This reinforced to some extent its position as a good candidate to take on new projects, but politically strategic maneuvering would always remain a primary requirement. Compromises remained necessary, which undermined the agency's 'can do' reputation.

The legitimacy mechanism helps explain the institutional rise of the PA. It also explains the development of a blind corner with regard to its political environment. Self-financing projects had always been the key to its legitimacy as an entrepreneurial agency. This financing principle was a justification for its existence and its expansion. It explains why the Authority would not listen to local demands for commuter rail improvement and to protests against the fourth jetport or the WTC. When the Authority was eventually forced to adopt the ailing rail system, it was criticized for organizational arrogance *and* financial mismanagement – which both damaged its legitimacy.

The PA does not seem to have developed the same rigidity as the TVA. Unlike the TVA, the PA had continuously been heavily dependent on political approval, forging coalitions between different stakeholders for each new project (see also Hargrove, 1994, p. 302). The PA had interpreted its mission in its broadest sense. By contrast, the TVA had narrowed its mission by primarily focusing on its role as a power supplier, and it aggressively pursued this single course. Once the legitimacy for its performance as a power supplier dropped, the TVA was left empty-handed: it had lost its distinct identity. The support base for the PA crumbled less rapidly, as it continued to book some infrastructural successes.

The PA's actions were not (entirely) driven by deterministic consequences of earlier choices. The Authority's de-institutionalization must be attributed to factors beyond its control: its heavy dependence on

political support and its legal impotence to define its own tasks. When new political actors took stage, big government became less popular, and railroad deficits were cast into its lap, maintaining high legitimacy became much harder.

The vigorous and unfortunate pursuit of nuclear energy by the TVA was not written in its legislative charters. It was driven by decisions made along the road of institutional development. The self-reinforcing spiral of supply and demand had increasingly defined its course. In addition, sunken costs of (nuclear) power facilities inhibited reversal of the organization's plans and change of its policy. Finally, the TVA's legitimacy and identity had come to rest on its ability to provide cheap electricity, which further drove the organization down its road to perdition.

Institutionalization can be reinterpreted as the active nurturing of path-dependence processes. It purposefully saps the abilities of both members and constituents to consider alternative ways of thinking and acting; it provides a forceful and legitimate argument for pursuing the course. Success thus guarantees more success, and institutions come into being. Path dependence is almost a condition for achieving success.

Maintaining success is a different matter, however. When established routines become counterproductive, they must be adapted. The organizations under study lost their flexibility to do just that. That does not have to be the case: 'proper' institutions find a balance between responsiveness and resilience, an 'institutional fit' with their environment – think of the U.S. Forest Service, the General Accounting Office or the Federal Bureau of Prisons (Kaufman, 1960; Walker, 1986; Boin, 2001). Too much success, perhaps gained too easily, disturbs this balance. It may well be that crises in the early stages of institutional development help to 'immunize' public organizations against rigidity and arrogance – a hypothesis that may inform future research on the topic.

Notes

* We found inspiration for this comparative design in Erwin Hargrove's suggestions for further research in his book Prisoners of Myth, 1994, p. 302, and in his comparison of NASA and the TVA in his 2004 conference paper (Leiden University, June 10–12, workshop 'Why Public Organizations Become Institutions') that can be derived from http://www.bestuurskunde.leidenuniv.nl/index.php3?&c=553#3. We are very grateful for Hargrove's invaluable comments in further email contact. Jameson Doig (2001) also alerts us to the similarities between both government agencies on the first page of his book.

1. De-institutionalization, and ultimately institutional demise, is here defined as the reverse of Selznick's institutionalization: the organization loses its

legitimacy and distinct identity in the eyes of its members and external stakeholders. The organization disintegrates into a collective enterprise to meet certain functional requirements characterized by a lack of shared understanding of mission and core values.

2. In fact, Mahoney (2000) distinguishes a functional perspective, a power resources perspective and a legitimizing perspective in addition to the efficiency perspective. Mahoney's presentation of the legitimizing and the functional perspective does not enable us to discern them conceptually in the case analysis, however. In this study, those two perspectives will be taken together as one mechanism because institutional legitimacy and functionality to its environment (meeting external expectations with institutional performance) seem to be intimately linked.
3. The entire account of the organizational development of the Tennessee Valley Authority is based on Erwin Hargrove (1994) Prisoners of Myth, University of Tennessee Press, unless otherwise mentioned.
4. The TVA had ambitiously set out to build *17* reactors after 1966.
5. 'By 1961 the federal share of TVA's budget was only a few million dollars because of the Bond Revenue Act of 1959. [...] The receding of a national mission was furthered by self-financing' (Hargrove, 1994, p. 153).
6. This account of the organizational development of the New York Port Authority is entirely based on Doig, J. (2001) Empire on the Hudson, Columbia University Press, unless otherwise mentioned.
7. 'Tobin called upon his associates to work as hard as he did, and he was impatient with staff delays in completing assignments; but he applauded their effort and successes, and he told the commissioners and the public, through the press, of their work. He learned the names of traffic officers, secretaries and other staff members in positions high and low throughout the organization, and he spoke with them – not just to them – as he visited offices, bridge crossings, and construction sites. In 1944, Tobin and the commissioners established Port Authority Service emblems to be worn by employees who had been with the Authority for five years or longer, a Medal of Honor for employees who had carried out meritorious acts at personal risk and the Distinguished Service Medal to recognize exceptional service of employees on the job' (Doig, 2001, p. 286; see also Walsh, 1978, pp. 224–226).

References

Alink, F., R.A. Boin and P. 't Hart (2001) 'Institutional crises and reforms in policy sectors: The case of refugee policy in Europe', *Journal of European Public Policy*, vol. 8 (2), pp. 286–306.

Boin, R.A. (2001) *Crafting Public Institutions*. Boulder: Lynne Rienner.

Boin, R.A. and P. 't Hart (2000) 'Institutional crises and reforms in policy sectors', in H. Wagenaar (ed.) *Government Institutions*. Dordrecht: Kluwer Academic Publishers, pp. 9–31.

Carter, C.A. (2008) 'Identifying causality in public institutional change: The adaptation of the national assembly for Wales to the European Union', *Public Administration*, vol. 86 (2), pp. 345–361.

Colignon, R.A. (1997) *Power Plays: Critical Events in the Institutionalization of the Tennessee Valley Authority*. Albany: SUNY Press.
Collins, J.C. and J.I. Porras (1994) *Built to Last: Successful Habits in Visionary Companies*. New York: Harper Business Publishing.
Doig, J.W. (2001) *Empire on the Hudson*. New York: Columbia University Press.
George, A.L. (1979) 'Case studies and theory development: The method of structured, focused comparison', in P. Lauren (ed.) *New Approaches in History, Theory and Policy*. New York: The Free Press, pp. 43–68.
Hargrove, E.C. (1994) *Prisoners of Myth: The Leadership of the Tennessee Valley Authority*. Knoxville: The University of Tennessee Press.
Hargrove, E.C. (2004) 'Historical comparisons', paper presented at the Leiden University Conference *Why Public Organizations Become Institutions*, Leiden, Netherlands, July 10–12, 2004.
Hargrove, E.C. and J.C. Glidewell (eds) (1990) *Impossible Jobs in Public Management*. Lawrence: University Press of Kansas.
Hay, C. (2001) 'The "crisis" of Keynesianism and the rise of neoliberalism in Britain: An ideational institutionalist approach', in J.L. Campbell and O.K. Pedersen (eds) *The Rise of Neoliberalism and Institutional Analysis*. Princeton: Princeton University Press, pp. 193–218.
Jervis, R. (1997) *System Effects: Complexity in Political and Social Life*. Princeton: Princeton University Press.
Kaufman, H. (1960) *The Forest Ranger: A Study in Administrative Behavior*. Washington, DC: Resources for the Future.
Kaufman, H. (1952) 'Gotham in the air age', in H. Stein (ed.) *Public Administration and Policy Development: A Case Book*. New York: Harcourt Brace and Co., pp. 143–198.
Kuipers, S.L. (2004) *Cast in Concrete? The Institutional Dynamics of Belgian and Dutch Social Policy Reform*. Leiden University: Dissertation.
Kuipers, S.L. (2006) *The Crisis Imperative. Crisis Rhetoric and Welfare State Change in Belgium and the Netherlands in the Early 1990s*. Amsterdam: Amsterdam University Press.
Landau, M. and D. Chisholm (1995) 'The arrogance of optimism: Notes on failure avoidance management', *Journal of Contingencies and Crisis Management*, vol. 3 (2), pp. 67–80.
Levitt, B. and J.G. March (1990) 'Chester I. Barnard and the intelligence of learning', in O.E. Williamson (ed.) *Organization Theory: From Chester Barnard to the Present and Beyond*. New York: Oxford University Press, pp. 11–37.
Mahoney, J. (2000) 'Path dependence in historical sociology', *Theory and Society*, vol. 29, pp. 507–548.
Miles, R.H. (1980) 'Findings and implications of organizational life cycle research', in J.R. Kimberley and R.H. Miles (eds) *The Organizational Life Cycle: Issues in the Creation, Transformation, and Decline of Organizations*. San Francisco: Jossey-Bass Publishers, pp. 430–450.
Pierson, P. (2004) *Politics in Time: History, Institutions, and Social Analysis*. Princeton: Princeton University Press.
Selznick, P. (1949) *TVA and the Grassroots: A Study of Politics and Organization*. Berkeley: University of California Press.
Selznick, P. (1957) *Leadership in Administration*. Berkeley: University of California Press.

Thompson, J.D. (1967) *Organizations in Action*. New York: McGraw-Hill.
Walker, W.E. (1986) *Changing Organizational Culture: Strategy, Structure, and Professionalism in the U.S. General Accounting Office*. Knoxville: University of Tennessee Press.
Walsh, A.M. (1978) *The Public's Business: The Politics and Practices of Government Corporations*. Cambridge: MIT Press.

Part III
Strategic Path Dependence

5
Path Dependence through Corporate Political Activity

Jan Siedentopp and Albrecht Söllner

5.1 Introduction

Corporate political activity (CPA) is as old as business itself. Pablo Spiller and Sanny Liao (2006) claim that the participation of interest groups in public policy making is unavoidable. Although this influence is met with some noteworthy suspicion from both policy makers and the public, it is generally assumed that 'lobbying' of firms and interest groups works in the favor of those who lobby.

Although this assumption is not being questioned in this chapter, we would like to point out that an over-concentration on CPA might also be problematic for firms since they might get locked into a strategy of government orientation (non-market strategy) that could prevent them from being customer oriented (market strategy) in the long run. Even though this might not be a problem when competitive pressure is low, firms might face economic problems when competitive pressure increases. A strategic lock-in with respect to government orientation makes shifting between market and non-market orientation a difficult task. By showing this path-dependent development, we hope to contribute to a better understanding of CPA and its effects on the strategic (in)flexibility and performance of firms.

In the following section we give a brief overview of the main concepts and ideas of the CPA literature. In Section 5.3 we develop a theoretical framework for deducing our hypotheses concerning a potential strategic inertia through CPA. Path-dependence theory and the resource-based view form the two pillars of the conceptual model. We then present an empirical test of our hypotheses and a discussion of the results. In the last section we present conclusions, including managerial implications and some thoughts about further research.

5.2 Corporate political activity – a brief introduction

CPA is defined as corporate attempts to shape government policy favorable to the firm (Baysinger, 1984). The main reasons for CPA are the firm's desire to change the rules of competition and to improve its market position. Since CPA seems to provide economic success, it has received increasing attention in recent years. David Baron (1995) is among those scholars who emphasize the importance of non-market influences on the firm's strategic development and performance. Baron (1995) proposes an integrated strategy of market and non-market components. Brian Shaffer (1995) discusses the consequences of public policies for the competitive environment of the firm and provides an explanation for political behavior. Whereas the public usually refers to CPA as 'lobbying', researchers are focusing on different types of CPA strategies and their potential impact on firm performance (Boddewyn/Brewer, 1994; Hillman, 2003; Hillman/Hitt, 1999; Meznar/Nigh, 1995). Although equipped with different perspectives, most work in this field converges to the common estimation that the engagement in CPA results in positive effects for firm performance (Hillman et al., 2004). Several empirical studies have demonstrated the positive impact of CPA on firm performance, although isolating the effect of CPA on performance is rather difficult (Hillman/Hitt, 1999; Marsh, 1998; Shaffer/Hillman, 2000; Shaffer et al., 2000). Others have shown that the benefits of different CPA strategies depend very much on the circumstances under which these strategies are chosen (Spiller/Liao, 2006).

Up to now, only a few studies have been directed on the potential negative implications of CPA. For instance, Bonardi (2004) dealt with CPA strategies of former monopolies which had a negative influence on their international growth performance. Frynas and Mellahi (2003) analyzed firm-specific political risks of transnational corporations. Siegel (2007) in his study of South Korean enterprises found that elite sociopolitical network ties to the regime in power can quickly change between liability and asset for firms which depend on political activities, for example, in cross-border strategic alliances. It is the purpose of this paper to further our understanding of the potential shortfalls of CPA. Although there might be positive effects of CPA for the firm, we argue that there are also cases where an over-concentration on CPA has negative effects on firm's flexibility and performance. We claim that there is a trade-off between CPA representing a non-market strategy, and customer satisfaction representing a market strategy. We argue that firms might get locked into a non-market strategy path

through CPA that could prevent them from being customer-oriented in the long run.

Just to give a brief example: The biggest German telecommunications company, Deutsche Telekom, shows that a firm can be very well organized and successful in its non-market strategies by ever-increasing engagement in CPA and influencing political decisions on telecommunication regulation in Berlin and Brussels. At the same time, however, their market strategy is not successful in raising their level of customer satisfaction. In addition, during the last few years, hundreds of thousands of customers have not renewed their contracts and moved to competitors. Why is that the case and, more interestingly, does their non-market strategy focus prevent them from achieving better market performance?

5.3 Theoretical framework

In this section we will use the resource-based view and path-dependence theory in order to develop a theoretical framework that addresses the above-mentioned research questions. In recent publications such as Dahan (2005a; 2005b) and McWilliams et al. (2002), the resource-based view offered a useful perspective on issues related to CPA, such as political resources. For our study, a resource-based perspective seems especially promising because increasing investments of resources in non-market strategies are to be seen as one of the core arguments for a potentially inefficient strategic path development. Since political resources are scarce, immobile and highly specific, they cannot easily be employed at the market strategy side of the company. Hence, these tangible and intangible corporate resources cannot be shifted into market strategies such as raising the customer satisfaction level of the company.

5.3.1 Political resources

Corporate resources are defined as '[...] any means of development controlled by a firm' (Dahan, 2005a, p. 10). Earlier works have already mentioned the need for mobilizing specific resources in order to influence public decisions successfully (Baron, 1995, p. 60; Boddewyn, 1993). Political resources therefore meet the definition of corporate resources. They are a means of development of the firm, by generating a favorable political-legal environment. In addition, the resources need not necessarily be owned by the firm as long as it controls them. A typology of political resources can be found in Dahan (Dahan, 2005b, p. 47), where he differentiates between *expertise, financial resources, relational resources,*

organizational resources, reputation among other non-market actors, public reputation, support of stakeholders, and *recreational skills*. Table 5.1 provides an overview of the various resource components.

However, in order to use the resource-based view in combination with any path-dependence argument, it would be necessary to measure investments in CPA-related resources. So far, indirect indicators such as spending on political action committees have been used in research to measure the level of CPA on a firm level. A conceptualization and operationalization of CPA as a theoretical construct has not yet been established in the literature (for initial research on this topic see Siedentopp, 2007a; Siedentopp, 2009). Therefore, one key challenge of this chapter is to operationalize the CPA construct on the basis of Dahan's (2005b) typology in

Table 5.1 Typology of political resources

Political Resource	**Components**
Expertise	Expertise in technical, managerial, social, ecological, legal, and political fields.
Financial resource	Campaign contributions (directly) and funding investments in other resources (indirectly) for political action.
Relational resource	Sitting on a public committee (formal relationship) or personal contacts (informal relationship).
Organizational resource	Government relations department, legal unit, etc. (internal) or hired consultancy office, trade association, etc. (external).
Reputation among non-market actors	Reputation of the firm's leaders or advocates (individual reputation) and the reputation of the firm itself (institutional reputation).
Public reputation	Short-term perception of the politically active organization by public opinion and the media.
Support of stakeholders	Legitimacy and political weight through: Coalition building (short term), strategic alliances (short to mid-term), business association membership (long term), number, unity and variety of the supporting stakeholders (formal to informal).
Recreational skill	All kinds of recreational services supplied to public decision makers and journalists in order to gain the opportunity for a more informal, personal and perhaps favorable contact. Includes diverse services like: wining and dining, free trips, as well as social events.

Source: According to Dahan (2005a, p. 14; 2005b, p. 47).

order to evaluate path dependence through CPA. The second main challenge is to prove the existence of path-dependence mechanisms that in the end lead to a strategic lock-in. Path-dependence theory is applied for this second challenge and its logic will now be described briefly.

5.3.2 Path-dependence theory

The general claim of path-dependence theory is that over time a (strategic) process can, at a final stage, lead to a lock-in situation, where change will be very hard, costly or even impossible (Arthur, 1989; 1994; David, 1985). Taking path-dependence theory as a rigorous perspective, it can be understood as an approach to explaining emerging process-related phenomena, thereby focusing on self-reinforcing mechanisms and forming specific strategic practices (a combination of routines and resources) which *can* eventually lead to a lock-in situation (Sydow et al., 2009). In other words, path dependence tries to explain how strategic scopes narrow down over time.

The most prominent example of path dependence was described by Paul David (1985). He explained why the QWERTY keyboard succeeded over its alternative and, in his opinion, more efficient keyboard formations. Although his reasoning drew opposition from economists like Liebowitz/Margolis (1990), path-dependence theory has since gained much support. More recently, path-dependence examples can be found in Burgelman's (2002) case study of the Intel Corporation, in which he explained the development of a strategic inertia situation within a firm, and in many other contributions (e.g., Black/Boal, 1994; Dougan/Bronson, 2003; Garud/Karnøe, 2001; Greener, 2002; Karim/Mitchell, 2000; Koch, 2006; Lamberg et al., 2004; Mahoney, 2000; Pierson, 2000; Siedentopp, 2008; Siedentopp, 2009; Wilts, 2006).

Six types of self-reinforcing mechanisms are held responsible for a process of path dependence: *economies of scale and scope* (main focus: cost), *direct and indirect network externalities* (utility), *learning* (skills/capabilities), *adaptive expectations and expectations of expectations* (standardization), *coordination effects* (interaction), and *complementary effects* (connectivity) (Koch, 2006). Figure 5.1 illustrates how a path-dependent process can be divided into three stages.

First, we find a phase of increasing selectivity, wherein the firm's management makes certain decisions. A second phase is marked by positive feedback that makes other behavioral options less and less attractive. It is the above mentioned self-reinforcing mechanisms that might in the end lead to a final stage of path dependence where it is extremely difficult for a company's management to change the once-chosen path. The

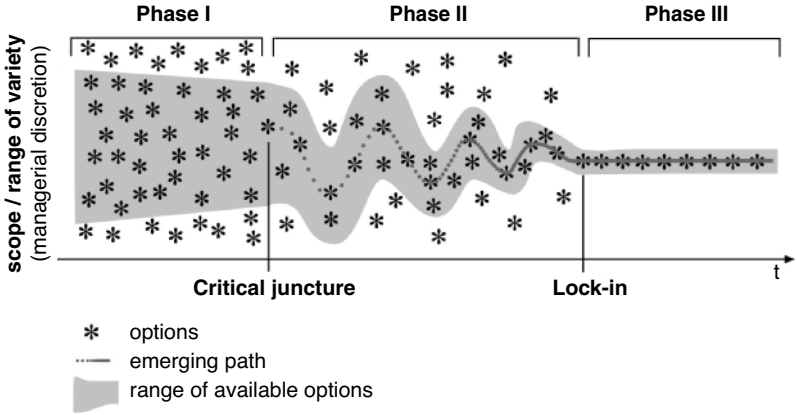


Figure 5.1 Model of a path dependence process
Source: Sydow et al. (2009, p. 692).

transition from phase I to phase II is often characterized by a critical juncture or some kind of critical event.

Within the CPA literature, path-dependence reasoning has rarely been used explicitly. Despite the attention given to the role of inertia in organizational and industry evolution (e.g., Hannan/Freeman, 1984) and although Lamberg et al. (2004) and Skippari (2005) looked at CPA from a process perspective, researchers have paid little attention to how inertia in a CPA path-dependent process evolves and may become a significant source of business disadvantage. An exception is Wilts (2006) who writes about path dependencies in corporate routines and actors' knowledgeability about CPA strategies. Generally, most studies used the notion of path dependence in a rather metaphorical way.

Our study addresses this gap. It seeks to shed light on the role of an increasing focus and resource shift towards CPA strategies that leads to strategic (non-market) inertia. Taking a closer look at the different political resources discussed by Dahan (2005b, p. 47) we are going to show that there are strong forces that keep firms on a CPA path in phase II of the model of a path-dependent process (see Figure 5.1) and prevent them from other strategic orientations.

5.3.3 Hypotheses

From our theoretical framework we derive two hypotheses:

HYPOTHESIS 1: A CPA strategy (non-market strategy) has a strong propensity to lead into a situation of path dependence and lock-in.

HYPOTHESIS 2: A CPA strategy (non-market strategy) lock-in will reduce the firm's focus on market strategy, for example, on customer satisfaction.

About Hypothesis 1) The reasons for a company to get involved in and stay on a CPA path (phase II) are manifold. Regulation plus issues such as taxes, labor, and environmental laws affect business. Therefore, management has to consider these political issues. Another reason companies get involved in CPA is competition. If competing companies within a sector get involved into CPA and build up their own political interest networks (*relational resources*), it is almost impossible for a company to opt-out or refrain from CPA. Consequently, CPA units are being built into the organization (*organizational resources*). Once a company has decided to build up a CPA unit and has it running for some time (*expertise* and *financial resource allocation*), drawing out of CPA would imply great sunk costs in terms of loss of reputation and image on the political and stakeholder side (*public reputation* and *reputation among other non-market actors*). The laborious path of building a company's political reputation is quickly lost with a decision for draw back. One of the core arguments used in the theory of path dependence seems applicable here. Many of the resources allocated for CPA have to be considered *specific* in the sense of Williamson (1975; 1985): Once deployed they are *sunk* and irretrievable. Building up a public affairs department not only leads to sunk costs due to the fact that the resources cannot just be transferred to another use; it is also the *public reputation* and most importantly the *relational resources*, for example, committee memberships and personal networks to political decision makers, that would be at stake in the case of closing down the department or terminating the relationship with administrative and government officials. *Learning* and building up political *expertise* can be seen as another strong source of self-reinforcement. North (1990, p. 103) as well as Denzau/North (1994) have emphasized the role of ideologies and mental models that may lead to institutional rigidity. Haase, Roedenbeck and Söllner (2007) claim that once a certain knowledge or ideology has been adopted, individuals tend to see the same problems and to use the same instruments to solve them. Accordingly, individuals can only define or identify new problems on the basis of the mental models already available to them. If individuals build up knowledge in fields such as politics, public administration, social issues, environmental issues, etc., it is much easier to extend this existing knowledge than to start learning in a completely new field. Consequently, if a CPA expertise worked perfectly well in the past, a firm will not see a need to change that process in the future. Even if a changing market environment was to make a strategic shift

necessary, the organizational culture of relying on CPA might prevent a firm from change.

About Hypothesis 2) Our second claim is that CPA as a non-market strategy works against a customer-focused market strategy. One argument in favor of this claim is very straightforward: From a resource-based point of view, resources that are used for a non-market strategy cannot be used for a market strategy. This not only holds true for resources such as *relational resources* or numbers of employees in Department A or B, it also applies to *expertise* as a resource. Expertise about the functioning of political systems and administrations are of limited help when it comes to understanding customer needs. Although one could argue that the number of employees in a public affairs department is relatively small compared to the total number of employees working for a firm, we claim that if the firm's management is more focused on non-market strategy than on market strategy, this will have an impact on most of the structures and processes of that firm. Since we assume a self-reinforcing mechanism regarding the CPA path, we also expect an increasingly negative impact of a rising CPA level on the firm's market strategy level, for example, customer satisfaction. In the end this might lead into the last stage of the path-dependence model – a non-market strategy lock-in situation.

The *Resource Gap Model* in Figure 5.2 visualizes the diversion of non-market and market strategy within a path-dependent process. The level

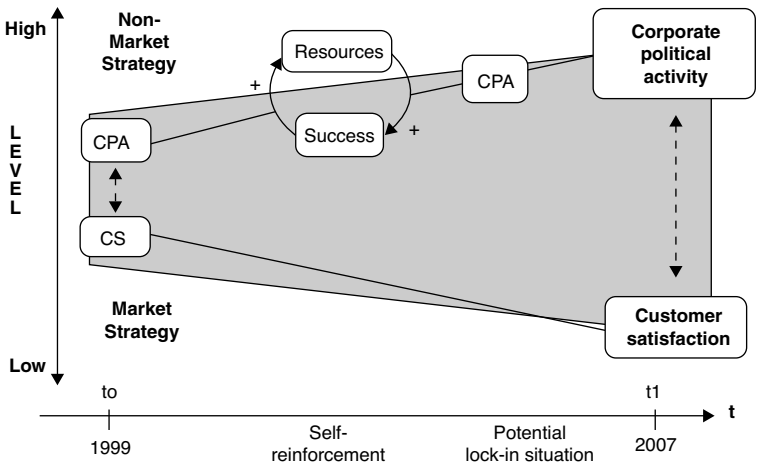


Figure 5.2 Resource gap model

Source: According to Siedentopp (2007a, p. 408).

of CPA on the non-market side and the level of customer satisfaction on the market side follow opposite directions. While CPA increases through self-reinforcing mechanisms, customer satisfaction decreases.

5.4 Empirical test and discussion of the results

5.4.1 Research design and data

In order to empirically evaluate the above hypotheses, it is necessary to answer two questions:

1. Has the level of CPA increased while the level of customer satisfaction mutually decreased during the period of time analyzed? If yes:
2. Have self-reinforcing mechanisms played a role in the path-dependent development of CPA?

To do so, the study follows a sequential explorative research design described by Cresswell (2003). First, in the explorative part, CPA was operationalized with the help of semi-structured interviews. The interviews, each lasting between 60 and 120 minutes, were conducted with over 40 German practitioners working within the field of Public Affairs (PA). A broad selection of interviewees was chosen to receive an adequate picture and idea of the whole PA field. Methodological emphasis, however, was placed on the quantitative part of the research design, which was conducted as a survey. The survey was sent out to 218 German companies in 2007 and addressed to people working as heads of public or government affairs departments in German companies. The sample companies had to be selected by self-research since no public register for public or government affairs activities exists for Germany. The selection criteria were that a company either had to have a separate public or government affairs unit or at least one person who was solely responsible for public and government affairs issues. The sample includes all DAX 100 companies, but also large-scale unlisted companies. The return rate of the survey was 47 per cent which makes this survey, to our knowledge, the most extensive one yet conducted in Germany focusing on the CPA of German companies. Some items within the survey asked for both current and past evaluation (turnover with public institutions, position to which to report CPA, number of employees working for PA issues, and PA budget), in order to collect data about the development of CPA over time.

The period of time covered in our analysis is from 1999 to 2007. Although determining a critical juncture, which triggers phase II of

the model presented in Figure 5.1, might not be an easy task in the case of CPA, a number of researchers and practitioners claim that the move of Germany's capital from Bonn to Berlin in 1999 marked such a critical juncture and that thereafter business-government relations have emerged into a new phase. The historically strong role of business associations in Bonn started to decrease and especially large-scale enterprises took their role in presenting themselves in the political arena of Berlin (Speth, 2006; Streeck et al., 2006). Furthermore, the time span of our study marks a time of deregulation (e.g. infrastructure, telecommunication) and increased competition due to globalization and fading trade barriers. Therefore, one could expect a shift from non-market (CPA) to market orientation (e.g., customer satisfaction) in many industries. If firms did not redirect their strategy and resources towards more customer and market orientation, this could be interpreted as an indicator for a path-dependent development of CPA. Self-reinforcing mechanisms might be at work.

For measuring market-strategy we used the company-specific data on the level of customer satisfaction that was provided by the *Servicebarometer AG, Munich* (2007). This index data is collected with standardized methods on a yearly basis. More than 20,000 customers are asked about their level of satisfaction with companies from most service and industry sectors. The question which creates an index for 'global customer satisfaction' is stated as follows: 'How satisfied are you with the service of this (mainly used) vendor?' The respondents were asked to reply on a five-point Likert scale ranging from fully satisfied (1) to unsatisfied (5).

5.4.2 Descriptive statistics

In order to answer the question of whether the level of CPA increased from 1999 to 2007 we will first present several descriptive statistics on the CPA development. Secondly, we show the correlation between the companies' CPA level over time and their respective level of customer satisfaction.

An indirect measure of overall CPA level development of German companies is whether and when companies have founded their own PA unit(s). The statistics indicate that the situation changed in 1999 when Berlin became the new capital of Germany. A large number of companies (54 per cent) founded their representative office in 1999 or later (see Figure 5.3).

Another indirect measure on overall CPA level development is presented in Table 5.2. It shows the development of the number of

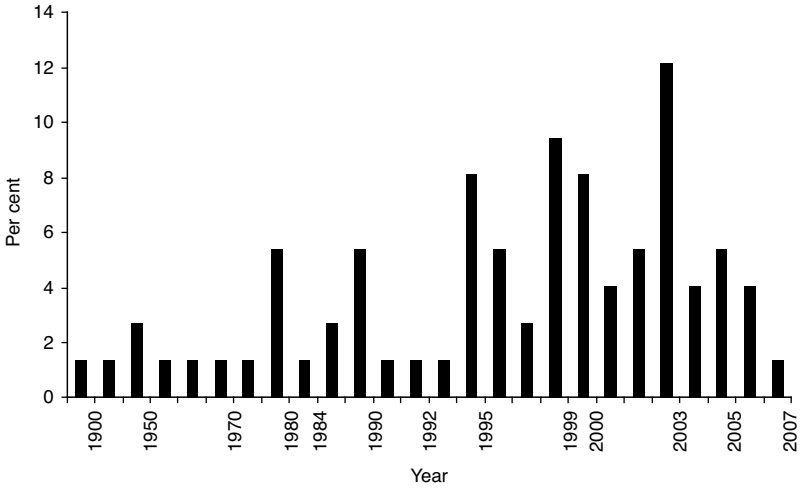


Figure 5.3 Founding year of public affairs units (in %; N=74)
Source: Siedentopp (2007b, p. 22).

Table 5.2 Development of PA employees: responsibility for Germany vs. EU (N=87)

	1999		2007	
	GER	**EU**	**GER**	**EU**
Mean	10.6	7.8	17.3	14.8
Median	3	1	6	4
Stand. Deviation	25.9	22.8	30.1	29.0
Max. Number	150	135	150	150
Total Employees	516	357	1256	957

Source: Siedentopp (2009, tables 40–42).

employees working in PA units. For Germany, in 1999 the mean number was 10.6 employees which increased to 17.3 in 2007; or by 63 per cent. On EU level, the PA employee number almost doubled from 7.8 in 1999 to 14.8 in 2007.

In order to support Hypothesis 1, it would be necessary (however, not yet sufficient) to show an increase in the foundation of PA units after 1999 and the number of PA employees. Indeed we find a noteworthy increase for both criteria. To get a third indication of an increase in CPA level, the figures in Table 5.3 show PA budget development since

Table 5.3 Public affairs budget development from 1999–2007 and 2008–2010

PA Budget 1999–2007	No.	Sampling Rate in %
Increase	47	54
Decrease	6	7
No change	34	39
PA Budget 2008–10		
Increase	31	35
Decrease	7	8
No change	51	58

Source: Siedentopp (2007b, p. 21).

1999 and future budget planning until 2010. The majority of companies (54 per cent) has increased their PA budget. This finding is supported by the fact that 16 per cent of the companies have increased their budgets since 1999 by 100 to 400 per cent. For the future development of PA budgets, 58 per cent say that they will not increase their budgets further. However, over one-third of the companies have made plans to increase their budget even further, indicating an ongoing investment in CPA.

Support for Hypothesis 2 can be found when looking at the correlation between the level of CPA and customer satisfaction for the time period 1999 to 2007. In order to calculate this correlation, we first formed an additive index with the above mentioned ex-post CPA items. Each item was coded with –1 (indicating an increase since 1999), 0 (no change), and +1 (a decrease since 1999). In total, we were able to correlate the CPA and customer satisfaction data for 21 companies. The result of the Spearman correlation coefficient was –0.418 with a 10 per cent level of significance (0.059; two-tailed test). As proposed by Hypothesis 2 we interpret this result as follows: A rising level of CPA between 1999 and 2007 is accompanied by a falling level of customer satisfaction within that period. Hence, a CPA strategy (non-market strategy) focus reduces the firm's focus on customer satisfaction (market strategy).

The following second empirical part explores in more detail, which self-reinforcing factors played a role in the observation of 'path-dependence through CPA' and if a sufficient explanation for Hypothesis 1 can be provided.

5.4.3 Self-reinforcement indicators for the CPA path

For the operationalization of the CPA construct and its influence on the latent construct of 'path dependence through CPA' we will use the

methodology of Partial Least Squares (PLS) and follow the suggestions of Chin (1998) and Lohmöller (1989). PLS is valued as a powerful method of analysis because of the minimal demands on measurement scales, sample size, and residual distributions (Wold, 1982). It can be used for theory confirmation, to suggest where relationships might or might not exist, and to suggest propositions for testing. Since the iterative algorithm performed in a PLS analysis generally consists of a series of ordinary least squares (OLS) analyses, identification is not a problem for recursive models nor does it presume any distributional form for measured variables. PLS starts with the goal to help the researcher obtain determinate values of the latent variables (LV) for predictive purposes.

Since PLS makes no distributional assumption, traditional parameter-based techniques for significance testing/evaluation would not be appropriate. Instead, Wold (1980; 1982) argued that PLS should apply prediction-oriented measures that are also non-parametric. Therefore, the R^2 for dependent LVs, the Stone-Geisser test Q^2 (Geisser, 1975; Stone, 1974; Wold, 1982) for predictive relevance of the model, and the effect size f^2, to see whether the impact of a particular independent LV on a dependent LV has substantive impact, are applied. The interpretation of the R^2 and the corresponding standardized path estimates is identical to that of traditional regression. The sample reuse technique Q^2 has been argued as fitting the soft modeling approach of PLS 'like hand in glove' (Wold, 1982, p. 30). $Q^2 > 0$ implies the model has predictive relevance, whereas $Q^2 < 0$ represents a lack of predictive relevance.

On the left side of the PLS model we used 14 items (see Appendix II) to measure three dimensions to represent the latent construct of CPA: *Degree of Political Activity*; *Degree of CPA Organization*; *Strategic Importance of CPA* (Siedentopp, 2009, p. 103 ff.). These dimensions were deduced from our theoretical framework. Since 'path dependence through CPA' itself is considered to be a latent construct, the aim was to find the adequate items that best represent and explain this construct. Five items, deduced from theory beforehand and cross-validated with the empirical results of the interview data, were chosen from the questionnaire. They are located at right side of the PLS model in Figure 5.4. Table 5.4 states the wording of these items.

Item 'Path01' indicates the historic dimension of CPA within a company, reflecting phase I of a path-dependent process (see Figure 5.1). Phase II is reflected by item 'Path02', in which CPA increases over time. It also shows the tendency to further increase CPA. All other items reflect phase III which indicates path dependence. All items used a 7-point Likert scale, where 1 meant 'do not agree at all' and 7 meant 'totally agree.'

Table 5.4 Items of the latent construct 'Path Dependence through CPA'

Short	Item	Phase
Path01	We often use a similar approach for our public affairs management.	Historicity
Path02	In the future, public affairs management will gain an increasing weight when it comes to strategic decisions for our company.	(Self-) Reinforcement
Path03	Frequently, in our core business we find mutual dependence between politics and our company.	Path-Dependence
Path04	The influence of political institutions constrains business decisions in our company.	Path-Dependence
Path05	Frequently, we find different objectives between politics and our company after applying corporate political activities.	Path-Dependence

In order to validate the measurement model, the first test is for multicollinearity between the items of the LV. The PLS model was calculated with the SmartPLS 2.0 (M3) software package (Ringle et al., 2007). After looking at the variance inflation factor (VIF) values for all items including the items for the exogenous LV, multicollinearity can be excluded (all VIF < 10). Secondly, indicator relevance is tested by first looking at the values of path weights for each indicator and secondly running a t-test for significance. Figure 5.4 shows the results of the structural model of path dependence through CPA and states the path weights and R^2 value. Only the path coefficient for the factor 'Degree of Political Activity' is not significant. For theoretic reasons, we decided to leave the factor included in the model (Lohmöller, 1989). Also, we find that all path coefficients have the predicted sign. Hence, Hypothesis 1 can be approved. Concerning the items of the endogenous LV, only the item 'Path04' does not comply with Lohmöller's critical value of 0.1 for the path coefficients (1989, p. 60 f.). The highest path weight can be found with item 'Path02', indicating the highest impact on the latent construct. The impact of CPA on future strategic business decisions is significantly high when it explains path dependence through CPA. This result goes along with the high weight of the dimension 'Strategic Importance of CPA' which has a significant weight of 0.749. In order to validate the structural model we first find the R^2 value to be exceptionally high with 0.692. This means that through this model we can

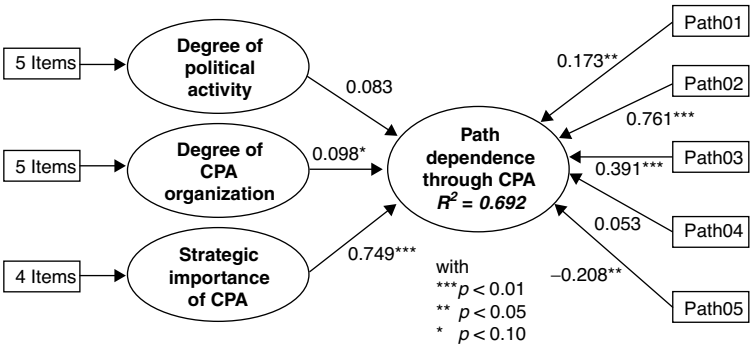


Figure 5.4 Structural model of path dependence through CPA
Source: Siedentopp (2008).

explain 69.2 per cent of the whole variance of the latent construct 'Path dependence through CPA'. Further, we look at the effect size f^2 of each factor and the Stone-Geisser criterion Q^2 for predictive relevance of the model. The dimension 'Degree of Political Activity' (0.02) and 'Degree of CPA Organization' (0.02) have a small effect on the LV. The dimension 'Strategic Importance of CPA' has the highest and most substantial effect on the endogenous LV indicated by an effect size of 1.36. The Q^2 value for the model is 0.158 indicating a reliable predictive relevance of the model. In other words, path dependence through CPA can be predicted through the impact of all three dimensions. Summarizing the results of the PLS model, Hypothesis 1 is approved by the PLS model. However, to support our argument that a CPA strategy has a propensity to lead into a situation of path dependence and lock-in, mechanisms of self-reinforcement must be identified.

Since self-reinforcement can have many underlying mechanisms, we chose to examine interview statements which describe how CPA can follow a self-reinforcing process and lead to strategic inertia. We evaluated five core interviews to identify the relevant aspects. Table 5.5 lists the relevant interview statements. We distinguish between a phase of self-reinforcement and a phase of strategic inertia in order to provide a better understanding of how and why CPA can narrow down a company's strategic options to a situation with inertial consequences for the overall strategy.

The statements highlight important strategic aspects of CPA for a company's overall business strategy. They show that CPA is closely linked to other operational business units and mark its mutual dependence.

Table 5.5 Indications of CPA self-reinforcement leading to strategic inertia

Phase	Statements
Self-reinforcement	1. '[...] especially for those companies, which strongly rely on regulation, e.g. pharmaceutical industry, energy sector [...] the **interplay between politics and managerial action play an ever major role**.' (Int. A)
	2. 'We (PA) are **closely linked to the operational units** of our company. We are asked for support on the political side for certain issues, e.g., regulation or sales. And, on the other hand, we ask them if they need support on the political frontier.' (Int. D)
	3. 'Traditionally business-government relations are very important to us.' (Int. D)
	4. '[...] We are **building up a false bottom**. Because, today the chances that you come a cropper with your business, in terms of image, are high; e.g., Coca-Cola.' (Int. B)
	5. '**Economy constantly has to work with politics**. E.g., issues like social reforms and labor market laws. [...] If the system is going through reforms, even more lobby work is done through the companies. [...] The whole **process is enforced through Europeanization and Globalization**.' (Int. E)
	6. 'Whole industry sectors are largely dependent on tax reallocation, e.g., agriculture, solar energy.' (Int. E)
	7. 'The experience shows that companies of a certain size, **need to permanently talk to politicians at higher ranks**. [...] If critical issues arise, you can rely on built-up trust and the political side has a basic understanding of your business.' (Int. C)
	8. A company '[...] can gain a better competitive position [...] not because it has made better managerial decisions, but because it has influenced the surrounding conditions towards its own interests.' (Int. E)
Strategic Inertia	9. '[...] if you want to do good marketing in the future, you have to convince the customer: Look, I have something with a value which exceeds the product itself. [...]Yes, this is something good or I am part of it. [...] Therefore, empty public affairs and PR campaigns are a farce. [...] then you really have a problem for years (...). **If you only miscalculated once, then you will struggle for years**; e.g., BP.' (Int. B)
	10. '[...] media perception of your action play an ever major role and in certain situations you only stick with decisions, because one agrees that it is not possible to present a total shift in strategy publicly. This would mean a major image, trust, and also a capital market loss. Internally one would possibly say: **Actually, we would decide differently, but this cannot be communicated, so we stick with the old decision**.' (Int. A)

Continued

Table 5.5 Continued

Phase	Statements
Strategic Inertia	11. 'Yes, **we do face strategic rigidities and low customer orientation through our tight corporate political activities**. [...] If we became less dependent on politics, we would enjoy greater strategic flexibility and higher market orientation.' (Int. D)
	12. 'A company can quickly be scuppered simply because it is closely monitored on what it does. Therefore, it has less freedom to act.' (Int. E)
	13. 'Those companies have a problem, which think that they can easily take business decisions; while de facto the rules of the game [...] are constantly changing. **The degree of freedom which you think you have is not as big as you might think**.' (Int. E)

Since CPA often serves as a preventive function towards potential negative future events for the company, it is often used for building a 'reputation buffer.' The other major goal which increases the strategic value of CPA is its role of influencing political decisions in favor of the company hoping to gain competitive advantage against others. As a consequence, CPA is in a constant process of interaction with political actors and other stakeholders. The inertial consequences which arise are multilayered. Short-winded PA and public relations campaigns can easily fall back on the company and impose long-lasting negative reputation among all stakeholders, including politicians. Strategic flexibility decreases dramatically once a certain strategy has been communicated among the stakeholders. Consequently, a company has to stick with old decisions, since a change of logic would impose a loss of trust and reputation. Overall, the degree of freedom to act and react decreases once political issues and actors need to be considered for managerial decisions on the market side.

5.5 Conclusion and implications

The negative consequences of the proposed strategic inertia through CPA at the company level described in this study are twofold: The main consequence can be seen in a low focus on the market strategy side. We showed a significant negative correlation between the level of CPA and customer satisfaction. In this respect, the genuine duty of a company to serve the customers and market needs to its satisfaction is

fading towards a minor priority. Resources and strategic direction cannot easily be shifted towards a market focus due to strategic inertia in the CPA focus of the company. Through the *historic evolution* of a firm, especially within former monopolistic markets such as infrastructure industries, and through *self-reinforcing* mechanisms, firms are vulnerable to losing their market focus. Thus, they fall back in competition if regulatory changes, such as the liberalization within the European Union, take place.

Taking a managerial perspective, implications can be found for the role of strategic decision makers within companies. The empirical results show that strategic factors of CPA have the highest effect on the construct 'Path dependence through CPA.' In other words, strategic decisions once made in favor or against CPA will have a significant influence on the future flexibility of the firm, whereas organizational and activity factors play a minor role. A look at the high weight of item 'Path03' (mutual dependence between politics and company) of the latent construct indicates that the executive management board might have to engage in reducing existing mutual dependency between politics and their company.

In terms of future research implications we call on focusing on the concept of Baron's (1995) market and non-market strategy and on the framework of path-dependence theory to explore how exactly CPA and market strategy can be beneficial for companies. For that purpose, structures and processes which are in place for both, market and non-market strategies, need further evaluation. A useful start could be to look at potential incentives which favor coordination efforts of both strategic directions. Overall, we make a plea for more empirical research in this area, especially from management researchers in Europe, to account for the strategic role of CPA today and to examine the opportunities and threats which CPA holds for businesses.

Appendix I

Interview A: General Manager, Public Affairs Consultancy, 14 December 2006
Interview B: General Manager, Public Private Partnership Consultancy, 23 November 2006
Interview C: Vice President Government Affairs, Company, 29 November 2006
Interview D: Head of Associations and Academic Relations, Company, 16 November 2006
Interview E: Academic Director, Scientific Institute, 15 November 2006

Appendix II

Table 5A.1 Items to measure the 'Degree of Public Affairs'

Dimension	Item
Degree of Political Activity	1. 'Please estimate how often, on average, informal talks with political actors are part of your task within your public affairs activities.'
	2. 'We analyze the law and regulation plans relevant to our company.'
	3. 'We have comprehensive knowledge about the positions of key actors which can influence political decision processes relevant to our business.'
	4. 'We influence the arrangement of the legal guidelines.'
	5. 'We analyze political and societal changes which influence our business.'
Degree of CPA Organization	6. 'We have installed an independent organizational unit in our company to represent our interests in the field of public affairs.'
	7. 'In order to be successful in our public affairs activities, an own representative office in the capital is not needed.' (reversed coded item)
	8. 'How many directly (e.g. in representative office) responsible employees for Germany work in your company in the field of public affairs?'
	9. 'In the case of a political issue which is highly relevant for our success, our management board invests personal effort at the political level.'
	10. 'Please estimate how often, on average, cooperation with public affairs agencies is part of your task within your public affairs activities.'
Strategic Importance of CPA	11. 'Today, public affairs management is the central aspect of the successful implementation of the overall business strategy.'
	12. 'The management board in our company has an understanding about political processes.'
	13. 'The majority of our approaches in the field of public affairs are future oriented.'
	14. 'The majority of our approaches in the field of public affairs are directed towards crisis management.'

Note

Both authors would like to express their gratitude to the German Research Foundation (DFG) for its substantial support of this research project; to the Servicebarometer AG, Munich for the provision of the customer satisfaction data; to the Fulbright Commission for supporting a sabbatical of Albrecht Söllner at the Haas School of Business at the University of California, Berkeley in 2008 where especially Pablo Spiller facilitated productive interactions with colleagues from different fields and provided a very thought-stimulating environment for further developing and discussing their ideas. An earlier version of this paper received valuable comments from six anonymous reviewers at the 70th Annual Meeting of the German Academic Association for Business Research (VHB), Berlin, May 2008, at the 19th Annual Meeting of the International Association for Business and Society (IABS), Tampere, June 2008, and at the Annual Academy of Management Meeting (AOM) 2008 in Anaheim.

References

Arthur, B.W. (1989) 'Competing technologies, increasing returns, and lock-in by historical events', *The Economic Journal*, vol. 99 (394), pp. 116–131.

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

Baron, D.P. (1995) 'Integrated strategy: Market and nonmarket components', *California Management Review*, vol. 39 (2), pp. 47–65.

Baysinger, B.D. (1984) 'Domain maintenance as an objective of business political activity: An expanded typology', *Academy of Management Review*, vol. 9 (2), p. 248.

Black, J. and K. Boal (1994) 'Strategic resources: Traits, configurations, and paths to sustainable competitive advantage', *Strategic Management Journal*, vol. 15, Special issue, pp. 131–148.

Boddewyn, J. (1993) 'Political resources and markets in international business: Beyond Porter's generic strategies', in A. Rugman and A. Verbeke (eds) *Research in Global Strategic Management*. Greenwich, CT: JAI, pp. 162–184.

Boddewyn, J. and T. Brewer (1994) 'International business political behavior: New theoretical directions', *Academy of Management Review*, vol. 19 (1), pp. 119–143.

Bonardi, J.-P. (2004) 'Global and political strategies in deregulated industries: The asymmetric behaviors of former monopolies', *Strategic Management Journal*, vol. 25 (2), pp. 101–120.

Burgelman, R.A. (2002) 'Strategy as vector and the inertia of coevolutionary lock-in', *Administrative Science Quarterly*, vol. 47 (2), pp. 325–357.

Chin, W.W. (1998) 'The partial least squares approach for structural equation modeling', in G. Marcoulides (ed.) *Modern Methods for Business Research*. Mahwah, NJ: Erlbaum, pp. 295–336.

Cresswell, J.W. (2003) *Research Design. Qualitative, Quantitative, and Mixed Methods Approaches*. Thousand Oaks: Sage.

Dahan, N. (2005a) 'Can there be a resource-based view of politics?' *International Studies of Management & Organization*, vol. 35 (2), pp. 8–27.

Dahan, N. (2005b) 'A contribution to the conceptualization of political resources utilized in corporate political action', *Journal of Public Affairs*, vol. 5 (1), pp. 43–54.
David, P.A. (1985) 'Clio and the economics of QWERTY', *The American Economic Review*, vol. 75 (2), pp. 332–337.
Denzau, A.T. and D.C. North (1994) 'Shared mental models: Ideologies and institutions', *Kyklos*, vol. 47 (1), pp. 3–31.
Dougan, W.L. and J.W. Bronson (2003) 'Suboptimal technology adoption: The case of computer reservation systems in the travel industry', *Journal of High Technology Management Research*, vol. 14 (2), pp. 289–305.
Frynas, J.G. and K. Mellahi (2003) 'Political risks as firm-specific (dis)advantages: Evidence on transnational oil firms in Nigeria', *Thunderbird International Business Review*, vol. 45 (5), pp. 541–565.
Garud, R. and P. Karnøe (2001) 'Path creation as a process of mindful deviation', in R. Garud and P. Karnøe (eds) *Path Dependence and Creation*. Mahwa, NJ: Earlbaum, pp. 1–38.
Geisser, S. (1975) 'The predictive sample reuse method with applications', *Journal of the American Statistical Association*, vol. 70 (350), p. 320.
Greener, I. (2002) 'Theorising path-dependency: How does history come to matter in organizations?' *Management Decision*, vol. 40 (5/6), pp. 614–619.
Haase, M., M. Roedenbeck and A. Söllner (2007) 'Institutional rigidity and the lock-in between mental models and ideologies', 11th Annual Conference of the International Society for New Institutional Economics (ISNIE) on *Comparative Institutional Analysis: Economics, Politics, and Law*, Reykjavik, Iceland, June 21–23, 2007.
Hannan, M.T. and J. Freeman (1984) 'Structural inertia and organizational change', *American Sociological Review*, vol. 49 (2), pp. 149–164.
Hillman, A.J. (2003) 'Determinants of political strategies in U.S. multinationals', *Business & Society*, vol. 42 (4), pp. 455–484.
Hillman, A.J. and M.A. Hitt (1999) 'Corporate political strategy formulation: A model of approach, participation, and strategy decisions', *Academy of Management Review*, vol. 24 (4), pp. 825–842.
Hillman, A J., G.D. Keim and D. Schuler (2004) 'Corporate political activity: A review and research agenda', *Journal of Management*, vol. 30 (6), pp. 837–857.
Karim, S. and W. Mitchell (2000) 'Path-dependent and path-breaking change: Reconfiguring business resources following business', *Strategic Management Journal*, vol. 21 (10/11), pp. 1061–1081.
Koch, J. (2006) 'Strategic responsiveness: A path dependency analysis of the German newspaper market of high quality journalism', EIASM Workshop on Organising Paths – *Paths of Organising*, Berlin, Germany, November 3–4.
Lamberg, J.-A., M. Skippari, J. Eloranta and S. Mäkinen (2004) 'The evolution of corporate political action: A framework for processual analysis', *Business & Society*, vol. 43 (4), pp. 335–365.
Liebowitz, S.J. and S.E. Margolis (1990) 'The fable of the keys', *Journal of Law and Economics*, vol. 33 (1), pp. 1–25.
Lohmöller, J.-B. (1989) *Latent Variable Path Modeling with Partial Least Squares*. Heidelberg: Physica.
Mahoney, J. (2000) 'Path dependence in historical sociology', *Theory and Society*, vol. 29 (4), pp. 507–548.

Marsh, S.J. (1998) 'Creating barriers for foreign competitors: A study of the impact of anti-dumping actions on the performance of U.S. firms', *Strategic Management Journal*, vol. 19 (1), pp. 25–37.

McWilliams, A., D.D. Van Fleet and K.D. Cory (2002) 'Raising rivals' costs through political strategy: An extension of resource-based theory', *Journal of Management Studies*, vol. 39 (5), pp. 707–723.

Meznar, M. and D. Nigh (1995) 'Buffer or bridge? Environmental and organizational determinants of public affairs activities in American firms', *Academy of Management Journal*, vol. 38 (4), pp. 975–996.

North, D.C. (1990) *Institutions, Institutional Change and Economic Performance*. Cambridge: Cambridge University Press.

Pierson, P. (2000) 'Increasing returns, path dependence, and the study of politics', *American Political Science Review*, vol. 94 (2), pp. 251–267.

Ringle, C.M., S. Wende and A. Will (2007) SmartPLS 2.0 (M3), http://www.smartpls.de (home page), date accessed 14 February 2007.

Servicebarometer AG (2007) Kundenorientierung: Inhalte, http://www.servicebarometer.com/kundenmonitor/km_studie/km_inhalte.jsp, date accessed 28 November 2007.

Shaffer, B. (1995) 'Firm-level responses to government regulation: Theoretical and research approaches', *Journal of Management*, vol. 21 (3), pp. 495–514.

Shaffer, B. and A.J. Hillman (2000) 'The development of business-government strategies by diversified firms', *Strategic Management Journal*, vol. 21 (2), pp. 175–190.

Shaffer, B., T.J. Quasney and C.M. Grimm (2000) 'Firm level performance implications of nonmarket actions', *Business & Society*, vol. 39 (2), pp. 126–143.

Siedentopp, J. (2007a) 'Development and measurement of corporate political activity: Indications of a path-dependent development', in B. Wempe and J.M. Logsdon (eds) *Proceedings of the 18th Annual Meeting of the International Association for Business and Society (IABS)*, Florence, Italy, May 31–June 3, 2007, pp. 406–411.

Siedentopp, J. (2007b) 'Public affairs management in Deutschland: Entwicklungen und perspektiven', *Working Paper No. 3*, Marketing Department Freie Universität Berlin, http://www.wiwiss.fu-berlin.de/institute/marketing/forschung/arbeitspapiere/index.html, date accessed 12 December 2007.

Siedentopp, J. (2008) 'Strategic inertia through corporate political activity: Consequences of a path-dependent process', *2008 Annual Meeting of the Academy of Management (AOM)*, Anaheim, USA, August 7–13, 2008.

Siedentopp, J. (2009) 'Public affairs-management von Großunternehmen: Markt- versus nichtmarktstrategien' (Public affairs management of large-scale enterprises: Market versus non-market strategies), Münster et al.: Lit.

Siegel, J. (2007) 'Contingent political capital and international alliances: Evidence from South Korea', *Administrative Science Quarterly*, vol. 52 (4), pp. 621–666.

Skippari, M. (2005) *Evolutionary Patterns in Corporate Political Activity: Insights from a Historical Single Case Study*, Doctoral Dissertation, Publication No. 522, Tampere: Tampere University of Technology.

Speth, R. (2006) 'Wege und entwicklungen der interessenspolitik', in T. Leif and R. Speth (eds) *Die fünfte Gewalt: Lobbyismus in Deutschland*. Wiesbaden: VS, pp. 38–52.

Spiller, P. and S. Liao (2006) 'Buy, lobby, or sue: Interest groups' participation in policy making – A selective survey', *Working Paper No. 12209*, National Bureau of Economic Research (NBER), Cambridge, http://www.nber.org/papers/w12209, date accessed 15 May 2008.

Stone, M. (1974) 'Cross-validatory choice and assessment of statistical predictions', *Journal of the Royal Statistical Society*, Series B, vol. 36 (2), pp. 111–147.

Streeck, W., J. Grote, V. Schneider and J. Visser (2006) *Governing Interests: Business Associations Facing Internationalization*. London and New York: Routledge.

Sydow, J., G. Schreyögg and J. Koch (2009) 'Organizational path dependence: Opening the black box', *Academy of Management Review,* vol. 34, pp. 689–709.

Williamson, O.E. (1975) *Markets and Hierarchies: Analysis and Antitrust Implications*. New York: The Free Press.

Williamson, O.E. (1985) *The Economic Institutions of Capitalism: Firms, Markets, Relational Contracting*. New York: The Free Press.

Wilts, A. (2006) 'Identities and preferences in corporate political strategizing', *Business & Society*, vol. 45 (4), pp. 441–463.

Wold, H. (1980) 'Model construction and evaluation when theoretic knowledge is scarce: Theory and application of partial least squares', in J. Kmenta and J.B. Ramsey (eds) *Evaluation of Econometric Models*. New York: Academic Press, pp. 47–74.

Wold, H. (1982) 'Soft modeling: The basic design and some extensions', in K. Jöreskog and H. Wold (eds) *Systems under Indirect Observation: Causality, Structure, Prediction (Part II)*. Amsterdam: North-Holland, pp. 1–54.

6

Continuity in Change – Path Dependence and Transformation in Two Swedish Multinationals

Olof Brunninge and Leif Melin

6.1 Introduction

Stability and change are central concepts in the field of strategic management. While companies are confronted with pressures for change, there are also inertial forces that tie organizations to established patterns. Hence, Melin (1998) describes the strategy process as a struggle between what exists and the possibilities for the future. Since the mid-1980s, a growing body of literature on path dependence has elaborated on the phenomenon of lock-ins during development trajectories (Arthur, 1989; David, 1985). Path-dependent developments are impossible or at least very hard to reverse and may result in inefficiency when necessary changes are prevented. Recently, Garud and Karnøe (2001) launched the alternative concept of path creation. While recognizing the existence of path-dependent developments, they emphasize the ability of entrepreneurs to make mindful deviations from established paths to accomplish change.

Usually, management literature depicts stability and change as a dichotomy (Mintzberg, 1978; Miller and Friesen, 1982; Tushman and Romanelli, 1985), meaning that at a given point in time, organizations and their strategies are neither stable nor changing. Garud and Karnøe (2001) take a more differentiated view, acknowledging that continuities can exist alongside changes. Changes even include historical resources that are used in a new configuration. However, also Garud and Karnøe's (2001) perspective puts a strong emphasis on disembedding from established structures.

In the present chapter, we present an alternative perspective on stability and change. We argue that the two can actually be combined in

processes of continuous development. In doing so, we see continuity as a special case of change, leading to novelty while avoiding radical ruptures from the past. The outcome of continuous change can always be recognized as the continuous follower of its antecedents, avoiding radical shifts from key characteristics of the past. Drawing on two cases of Swedish multinationals, the aim of this chapter is to investigate how change and stability are combined in strategy processes. We put particular emphasis on how managers make sense of change in those processes.

The empirical material for this chapter comes from extensive case studies on strategy processes in two large Swedish firms: the world's most profitable truck maker Scania and the bank Handelsbanken, being among the top five banks in Scandinavia (Brunninge, 2005). Data was collected primarily through 79 interviews with strategic actors at both companies. Interview data was complemented with extensive archival studies and close participant observation at events where strategic issues were presented and discussed. We used the data to construct organizational biographies (Kimberly and Bouchikhi, 1995) putting specific emphasis on the development of key strategic episodes (Hendry and Seidl, 2003) occurring during the period 1990–1995. By definition, strategic change only becomes visible in processual analysis. Our longitudinal approach allows us to describe and to analyze the interplay of stability and change along two key strategic episodes in Scania and Handelsbanken. Before going into the details of the empirical study, we will have a look at different views on stability and change in the strategy literature.

6.2 Views on stability and change in strategy research

While both stability and change are key concepts in management research, strategy literature often puts special emphasis on change, depicting it as a necessity for success that needs to be promoted by managers (Barr et al., 1992; Grinyer et al., 1988). Stability, on the other hand, is often seen as a problem that needs to be overcome (Reger et al., 1994). The rationale behind this picture is relatively simple: at least since Burns and Stalker (1961), we know that organizations, being open systems, need to realign themselves with turbulent environments. Hence, if organizations fail to change, this will eventually result in strategic drift (Johnson, 1987), meaning that they lag behind a changing environment and run into serious problems. However, in practice, change is often easier said than done. Hannan and Freeman (1977) note that

organizations often suffer from structural inertia. It implies that organizations have a tendency to resist changes as they reproduce established structures and strategies. This is partly related to internal normative agreements, but can also be attributed to external stabilizing forces, such as institutional pressures (DiMaggio and Powell, 1983). One major reason for inertia is cognitive, meaning that organizations and their members get stuck into established systems of meaning, like organizational identity (Reger et al., 1994) or organizational culture (Gagliardi, 1986; Johnson, 1987).

While the literature on organizational culture emphasizes cognitive reasons to inertia, the resource-based view of the firm (RBV, e.g., Wernerfelt, 1984; Barney, 1991) and the dynamic capabilities view (DCV, e.g., Teece et al., 1997) stress reasons that are linked to the resource endowments and routines in a firm. The RBV basically regards firms as bundles of resources (Wernerfelt, 1984). These resources have been created under firm-specific historical conditions, meaning that they can neither be easily replicated by competitors nor easily changed by the firm itself (Barney, 1991). Hence, while resources may be sources of competitive advantage over long periods, they also result in rigidities that may prevent firms from aligning itself with a changing environment (Anderson and Tushman, 1990; Leonard-Barton, 1992). Proponents of the DCV have emphasized how firms can adapt, integrate, and reconfigure their existing resources in order to sustain their competitive advantage over time (Teece et al., 1997). However, also dynamic capabilities in themselves develop along historical paths and changes are seldom frame-breaking (Eisenhardt and Martin, 2000; Teece et al., 1997). Also, the RBV and DCV literature addresses cognitive reasons of inertia. Not only can cognitive systems as organizational culture in themselves be seen as resources (Barney, 1986), they also influence how a firm interprets, uses and reconfigures its resources.

The concept of path dependence that was addressed in the introduction to this chapter has been added to the discussion on stability and change in strategy relatively recently. Addressing specific forms of inertia, one of its major strengths lies in its emphasis on the often neglected importance of history for organizational choice (Schreyögg et al., 2003). Compared to the path-dependence literature produced by institutional economists (Arthur, 1989; David, 1985), management literature using the concept puts more emphasis on managerial agency in path-dependent processes. For instance, Garud and Karnøe (2001) open up for path creation by entrepreneurs who make mindful deviations from established paths, while Schreyögg et al. (2003) point at the possibility

of de-locking path dependencies. Nevertheless, these authors affirm that it is difficult to escape from a path once it has been locked-in.

Managers hence face a difficult task: they need to constantly adapt their organization to a changing environment while at the same time struggling with counteracting forces that stabilize the firm. These forces may be both internal and external. Sometimes they even include the managers' own mindsets that seek to preserve existing structures and strategies. The issue of how to cope with the pressures for and against change becomes one of the major managerial challenges and involves critical questions: Do organizations have to go for either change or stability? Can the two be combined and if so, how can this be done?

Typically, management literature depicts the stability/change issue as an either-or question. Phases of stability in a firm's strategic development are supposed to alternate with phases of change that may be either radical (Miller and Friesen, 1982; Tushman and Romanelli, 1985) or incremental (Mintzberg, 1978). Such a view is problematic insofar that little room is left for stability and change occurring simultaneously. Strategies are complex, comprising a variety of different elements. It is not evident that all different aspects of strategy should either change or be stable at the same time. This implies that change and stability may actually be complementary, going hand in hand rather than excluding each other. The question of stability and change is further complicated by the fact that the labeling of a certain course of action as change is always a matter of interpretation. It is of course possible to measure change along pre-specified dimensions from the perspective of an external observer. Still, to the people involved in a strategy process, the perception of a development as stability or as change as well as of the magnitude of perceived change is the outcome of an interpretive process. It is critical for the strategy dynamics in an organization if and if so, how strategic developments are perceived as changes.

As indicated previously, the literature on strategy has put special emphasis on change while seeing stability as the potentially problematic absence of change in organizations. While stability thus seems to be easy to define, change can occur in different manners. The most common categorizations refer to the magnitude of change, using dichotomies like evolutionary and revolutionary change (Johnson, 1987). Tushman and Anderson (1986) even further subdivide radical changes, referring to changes in technology that can either be competence enhancing, when drawing upon the existing set of competences the organization has, or competence destroying when the established competences are rendered obsolete. When it comes to the sequence

of evolutionary and revolutionary changes, some scholars stress the virtues of small, incremental changes avoiding painful strategic revolutions (Quinn, 1978). Tushman and Romanelli (1985), on the other hand, propose a punctuated equilibrium model of change, claiming that organizations are characterized by long periods of relative stability that are disrupted by revolutionary changes. Reger et al. (1994), taking a normative stance, argue for middle-ground or 'tectonic' change avoiding the drawbacks of its evolutionary and revolutionary counterparts. Gebert (2000) criticizes the punctuated equilibrium model from a more fundamental point of departure. He notes that transformational and incremental changes are not necessarily opposed to each other as claimed by Tushman and Romanelli (1985). Rather, transformational change can also be the result of long-term, incremental changes that add up to a major transformation over time. In addition, even periods of incremental change are characterized by dynamism as seemingly stable situations often conceal tensions between different forces for and against change (Gebert, 2000). Also, Melin and Hellgren (1994) are critical of the stability/change dichotomy; they believe that a focus on magnitude is insufficient to characterize different types of change and propose to add the proactive or reactive nature as a second dimension to the study of strategic changes.

Some researchers actually discuss situations where change and stability either occur simultaneously or where changes are hidden behind a façade of stability. Gioia et al. (2000) as well as Chreim (2005) note that organizational features can change while the labels denoting these features remain stable. Strategic actors in an organization may use the stability in labels to conceal the changes they undertake, preventing resistance by organizational members. Already in 1957, Selznick remarked how organizational resistance can inhibit change. Introducing the concept of *institutional integrity,* he noted that over time organizations are infused with value for their own sake – they become institutionalized. Institutional integrity is only maintained as long as the organization's basic character remains intact. This does not rule out change *per se*. However, it implies that organizations do not tolerate change that is disruptive in relation to their basic character. There are examples where stability in some aspects of the organization has made it easier to change others (Hatum, 2002). Chreim (2005) makes the important point that change and stability, or continuity as she calls it, are a duality rather than two mutually exclusive opposites. This means that stability and change, though seemingly contradictory, are actually interwoven. One interesting, yet underexploited concept in the context of change that

may be helpful in dealing with the stability/change duality is that of continuity. In everyday language as well as in academic literature (e.g., Chreim, 2005) it is often used synonymously with stability. This use is misleading though. Continuity can denote change; however, a particular sort of change that occurs in non-disruptive manner. The change involved in continuity is not necessarily small, but it happens without ruptures that disconnect the new situation from the past. The outcome of change has enough similarity with history to be recognized as its continuous follower. Whether organizational members see a change process as continuous or not is of course a question of their subjective judgment. In this sense, a development is not continuous or discontinuous *per se*. Continuity is ascribed to a process by those interpreting it. This means that the question of whether strategy is developing continuously or not can be discussed. It also means that continuity, or discontinuity, can be constructed by pointing at things that either support, or contradict, the image of a continuous development.

6.3 The empirical study

Scania and Handelsbanken are two highly successful, internationally operating Swedish firms. Both companies were founded in the late nineteenth century and during the last decade they have been among the most profitable firms in their respective industries. Scania, having its roots in the manufacture of railway wagons, bicycles and cars, soon came to concentrate on trucks and buses. While buses accounted for the major part of Scania's turnover during some periods before World War II, trucks replaced them as the major product shortly after the war. Buses as well as engines that were sold separately, more and more became by-products of the increasingly successful truck business. Due to the small domestic market, Scania internationalized and gained strong footholds in Europe and Latin America. One important factor behind Scania's success was the company's product development philosophy, the so-called *modular system*. The system implies that the components in a Scania truck have standardized interfaces, allowing them to be combined in an almost indefinite number of different ways. Hence, the company can produce a wide variety of trucks with 'box of bricks' comprising a limited number of components. Thanks to modularization, Scania is able to realize significant economies of scale although it produces fewer units than some of its competitors. While the modular system is one key to Scania's success modularization also imposes restrictions to Scania's strategic options, making it difficult to move into certain markets and

segments. In North America, for instance, trucks are typically assembled from components produced by different firms. As Scania's modularization builds on the idea that all major components are made by Scania, the company has refrained from entering the North American market. Apart from this exception, Scania is today selling trucks in all parts of the world.

Handelsbanken was founded in 1871 as Stockholms Handelsbank. The bank subsequently acquired competitors all over the country and changed its name to Svenska Handelsbanken, or just Handelsbanken in 1921. Already at that time, Handelsbanken had established itself as one of the major banks in the Swedish market. Expansion abroad was, however, limited to serving the foreign business of Swedish companies as well as foreign firms doing business in Sweden. A highly regulated market provided a stable environment for Swedish banks. Despite the general stability, Handelsbanken ran into trouble in the late 1960s when profits went down and the bank was accused of violating Sweden's strict currency regulations. As a consequence, the Managing Director was dismissed and replaced by Jan Wallander, an economist who had successfully run a medium-sized regional bank for some years. Wallander initiated a radical decentralization program at Handelsbanken. Budgets were abolished and replaced with a benchmarking system. Central functions, including the central marketing department, were closed down. Decision-making authority was largely transferred to the local branches, which were supposed to be the most important organizational units in the decentralized organization. Wallander's turnaround strategy was successful and Handelsbanken quickly regained profitability. Decentralization was continued during Wallander's tenure and his successors have further developed the successful concept of branch-based banking. As the European financial sector was deregulated during the 1980s and 1990s, Handelsbanken was able to start serving local customers in foreign markets, starting with the neighboring Nordic countries. The idea of having decentralized operations with strong local branches was transferred to the new markets. Today, Handelsbanken has more than 600 branches in Sweden and abroad and has been among the most profitable Scandinavian banks since the early 1970s.

Scania as well as Handelsbanken can look back on a history of more than 100 years. Both companies are conservative in the sense that some of their characteristics remain relatively stable over long periods of time. At the same time, the companies have undergone significant changes since World War II. Interestingly, these changes have implied that each of the companies has developed a distinctive philosophy that is central

to operations: the modular system in Scania's case and the concept of decentralized banking at Handelsbanken. The two philosophies have been developed over several decades and have become key features of the companies' organizational identities. The strategy processes in both companies are characterized by the need to relate any strategic action to the respective management philosophy that is to keep their institutional integrity. While doing this, sometimes tensions arise as the need for external alignment may seem to call for more radical change. The way the companies deal with the dilemma of facing a changing environment and preserving their established ways of doing business will be illustrated in the following case episodes.

6.3.1 Case episode 1: Handelsbanken's introduction of internet banking

The emergence of internet banking in the 1990s constituted a major technology change in the banking industry. Most banks cherished the development as it was an opportunity to downsize their costly branch networks. For Handelsbanken the situation was different. How should a bank that claimed to be based on local branches, relate to a technology that seemed to render physical branch offices obsolete? Handelsbanken had maintained a relatively large network of branches in Sweden, while its competitors had been closing down and merging branches during the 1980s and 1990s. The bank claimed that local presence and decentralized decision-making were keys to its success. Now it seemed that the internet made geographical distance negligible. At least in principle, a wide range of services could be provided to an entire country from a single web branch. The remaining personal services might be concentrated to a limited number of branches in the major cities.

People at Handelsbanken, not least the powerful branch managers, were concerned that the new technology would mean the end of the bank's decentralization strategy. Not only would many branches be closed down, the autonomy of the remaining branches would also be reduced dramatically with a technology that allowed for centralized distribution of financial services.

Handelsbanken managers were thus hesitant to introduce a technology that might question the bank's basic business philosophy. On the other hand, refraining from internet banking was not a feasible option either. Handelsbanken was initially lagging behind when other Swedish banks started introducing internet banking. The solutions suggested by various consultants were too centralized to be legitimate in the organization. Finally, the bank adopted a solution that was radically

different from those of its competitors. Rather than having a centralized internet branch besides the traditional branch network, each of the more than 500 branches got its own website. Customers had to choose their branch when entering the website for the first time. Then the system stored a cookie on the customer's computer, directing him or her to the right branch the next time using internet banking. All local websites had the same graphical profile and shared all of the basic transaction functions. However, the branches were able to add local information, include tailor-made services for the local market and decide whether specific services should be actively promoted or not. Hence the branch managers retained their sovereignty over the bank's activity in each local market. Handelsbanken claimed that its internet solution was a new entrance to local branches rather than being a new distribution channel bypassing the branch network. The bank had been able to go through a major technological change while maintaining its established business philosophy.

6.3.2 Case episode 2: Scania's strategy as a producer of heavy trucks

During the last few decades, Scania has pursued a very focused strategy of making heavy vehicles only. Heavy vehicles, defined by Scania as trucks and buses over 16 tons, constitute the most profitable segment of the market. The focus on heavy trucks and buses is closely linked to the company's modular design philosophy. Offering trucks and buses over 16 tons, as well as industrial and marine engines that are derived from engines for those vehicles, implies that the set of components used by Scania is limited to this product range. Although the 'box of bricks' allows great variation, the possibilities are not unlimited. If Scania went for smaller trucks, the cabins would, for example, at some point become disproportional compared to the rest of the vehicle. Similarly, the current engine components would not necessarily be suitable for smaller engines. Hence, Scania's modular system restricts the strategic options of the company. Abandoning the strategy of focusing on heavy trucks would at some point mean that Scania would have to either make a significantly larger number of components or buy components – for example, smaller engines – from external suppliers. This would again make it difficult to capitalize on the advantages of modularization.

Over the years, Scania has faced pressure from the market to move into the medium-sized segment several times. Customers would appreciate buying trucks of different sizes from the same truck dealer, a demand that Scania dealers cannot satisfy with Scania trucks today. During the

last few decades, Scania has made several serious investigations on producing medium-sized trucks. The reason to finally dismiss the projects was always the modular system. Most recently, the issue was on the agenda at the end of the 1990s. At that time, some Western European countries thought of banning heavy traffic in city centers. This would have drastically limited the possibilities of using Scania's existing truck range for distribution purposes. Scania once again decided to launch an investigation into a medium-sized truck. Although the project went very far, it was finally discontinued as the restrictions in city centers were never introduced. Since the expected pressure from legislation did not materialize, Scania was happy to continue relying on the existing modular system. The question remains of how Scania would have rationalized producing a medium-sized truck if the company had been forced to do so. After all, Scania had for long time emphasized the benefits of its strategy to concentrate on the heavy-truck segment. A likely course of action would have been not only to stretch the modular system, but also the heavy-truck concept as such:

> The [change in strategy] would have been marginal. It would still have been the heavier end [of the truck market]. Just because some people say that heavy trucks are those that weigh more than 16 tons, you cannot say that over 12 tons is a light truck. Of course [you can say that they are heavy, too.] It's a matter of definition. Some countries have a threshold at 14 tons and some at 15. It would not have been a problem, but it never became an urgent issue. Chief Human Resource Officer (Interview 2003)

Scania was thus willing to make compromises regarding its modular philosophy and the concentration on the heavy segment if the firm would have been forced by the market. However, the company wanted to avoid declaring openly that this would actually be a significant change in strategy. By redefining heavy trucks rather than admitting that the company would produce medium-sized trucks, the company wanted to maintain the image that the successful, historically established strategy was still in place.

6.4 Discussion and conclusions

Both Scania and Handelsbanken are companies characterized by management philosophies that are historically established in the organizations and which have become key features of their organizational identities

(cf. Albert and Whetten, 1985). The two companies have also been highly successful since the management philosophies were introduced. The historical embeddedness of decentralization at Handelsbanken and modularization at Scania together with the success associated with these approaches, have contributed to institutionalizing the philosophies. For employees as well as many external observers Scania would be more or less unthinkable without modularization just as Handelsbanken would be unthinkable without decentralization. To use the vocabulary of the path-dependence literature, a lock-in of the philosophies has occurred as a result of the increasing returns they have generated. The two philosophies have thus become infused with value. They are cherished not only for their functional value – that is, their contribution to economic success – but also as ends in themselves, providing the organizations with a distinctive identity. Selznick (1957) addresses such institutionalization processes when discussing institutional integrity. Once characteristics of the organization have become institutionalized, it becomes a key managerial task to preserve these features in order to maintain institutional integrity.

For both Scania and Handelsbanken, path dependence is related to cognitive reasons. Abandoning the locked-in management philosophies would simply not be considered acceptable. Particularly in the Scania case, also a technological lock-in can be observed as the range of components is optimized for a specific product range. Any attempt to stretch the product range would not only require a major legitimization effort, it would also necessitate major investments in new components.

Our case episodes show that managers at Handelsbanken and Scania undertook significant efforts in order to protect decentralization and modularization, at least on the surface. Abandoning the philosophies would have threatened institutional integrity, probably met resistance among employees and other stakeholders, and put the trustworthiness of management into question. At the same time, both companies faced strong pressures from the environment to actually undertake such changes. The philosophies that once created increasing returns had potentially become dysfunctional. In Handelsbanken's case it was a technological shift in the banking industry that seemed to inevitably render the established way of distributing financial services obsolete. In Scania's case legislative pressures threatened to take away a significant part of the company's market if it decided to maintain its established strategy. For both companies there were at the same time strong internal forces pushing for stability and external forces pushing for change. The solutions the companies found to these dilemmas show that they were

not resistant to change *per se*. Both firms felt that it was (or in Scania's case was likely to become) necessary to change. However, the most obvious routes to change seemed locked as they appeared being threats to institutional integrity. Instead alternative routes were chosen.

Handelsbanken solved the tricky situation by using the historically established strategy with the decentralized approach as a template for its internet banking strategy. The approach to internet banking had to be chosen from the perspective of the branch-based bank, meaning that the branches needed to retain control of their customer relationships and the business associated with each customer. To customers it needed to be clear that the internet was just an additional access point to their local branch. It was their branch they visited on the web – not a separate internet branch or a central unit of the bank. Hence, the decentralized branch-based strategy that was crucial to Handelsbanken's identity remained stable. The aspects that changed were not critical from an identity point of view although the introduction of a new technology represented a significant change in practice. By using the old strategy as a template, Handelsbanken made the new strategy appear as the continuous follower of the old one. There was no rupture between the branch-based internet approach and the historically grown concept of branch-based banking.

Scania thought of creating a sense of continuity by redefining the meaning of heavy trucks. On the surface, the company would still have been a producer of heavy vehicles only. However, within the stable label 'heavy truck,' a new meaning was hidden. Instead of sticking to the established definition of more than 16 tons, Scania was willing to reduce the threshold to 14 or even 12 tons. This would also have meant revising the view of Scania's history. Whereas historical attempts to produce medium-sized trucks had traditionally been used as a terrible warning not to abandon concentration on heavy vehicles, they had now almost become historical forerunners to the new 'heavy' Scania.

The case episodes from Handelsbanken and Scania show how managers struggle to avoid changes in their companies that could threaten the institutional integrity of their organization. However, rather than avoiding change as such, they strive to reconcile the forces for and against change, by initiating changes that are perceived as continuous. Establishing continuity means that the institutional integrity (Selznick, 1957) of the firm is maintained and that changes that nevertheless occur are more likely to be accepted by organizational members. Where continuity is established, hence also the old identity of the firm is reinforced. This means that paradoxically even strategic

changes in an organization can have a stabilizing effect on the strategy process. There are several strategies for managers to establish continuity. These strategies are not mutually exclusive and can be combined in various ways:

- Using past strategies as a template for new strategies. The new strategies are formed in a way that is in line with the core elements of old strategies, such as when Handelsbanken based its approach to using the internet on its concept of branch-based banking.
- Adapting views of the past in order to fit to new strategies. The shared view of the organization's history is revised in order to legitimize strategic changes and to create a sense of continuity in strategic development, like when Scania thought of revising its view on former medium-sized trucks from examples of failures to examples of pioneering projects.
- Maintaining stable labels that are given changing meanings. The stability in labels conceals changes in strategy that actually occur (cf. Lindell et al., 1998; Gioia et al., 2000; Chreim, 2005), such as when Scania thought of changing its definition of heavy trucks in order to claim that it was maintaining its focus on heavy vehicles.
- Some elements of strategy are changed while those elements of strategy that are crucial to maintaining the organizations' identity remain stable, such as Handelsbanken's decentralization philosophy that remained untouched while a significant technological change occurred through the introduction of the internet.

The continuity established by such strategies lies in the eye of the beholder. It is a challenging task for managers to decide what changes can be undertaken and how these changes can be rationalized without threatening the institutional integrity of the organization. In this sense, managers become creators of continuity. If strategic developments are not continuous or discontinuous *per se,* continuity needs to be constructed by creating a shared picture of how new strategies relate to the organization's history.

With regard to path dependence this implies that the determinism lying in historically established paths creates fewer restrictions for future strategies than one might think at first sight. History does matter, but at the same time history is subject to interpretation and re-interpretation. When conceptions of history can be re-interpreted in order to fit better to future strategic plans or when historical labels are re-used with new meanings, organizations can escape from path

dependencies. This escape is, however, not experienced as a deviation like it typically would be when creating a new path in the sense of Garud and Karnøe (2001). While apparently following an established path, strategy actually takes a new direction. Change occurs, but it is perceived as continuous by organizational members. Similarly, even change that follows established paths can imply major transformations for a firm. Companies usually possess various resources that create different path dependencies. In the case of Handelsbanken, one such path dependence relates to the decentralized organization and the willingness to remain decentralized, while another relates to the traditionally trained bank staff and the resulting focus on transaction processing. While the decentralized organization is strongly infused with value, this is not true for the traditional training of staff to the same extent. Consequently, change is regarded as continuous as long as the institutionalized path of decentralized banking is followed. Deviations from the path of traditional transaction-based banking are much less significant for members' experience of continuity.

In this chapter, we have advocated complementing the concepts of stability and change with the concept of continuity, defined as a specific form of change. Our case episodes reveal that the stability/change dichotomy – which is often depicted in the strategy literature – is overly simplistic. In strategy processes, change and stability usually occur simultaneously. Even in situations of overall change, various elements of strategy may remain stable enough to make new strategies appear as the continuous followers of the old ones. Seeing continuity as a socially constructed phenomenon rather than an 'objective' feature of a change process, this implies that path dependencies relating to cognitive inertia (e.g., Reger et al., 1994) may sometimes be easier to overcome than it appears at first sight. Managers have various means at their disposal to make significant changes appear as being continuous in relation to past strategies. This does not mean that organizations will accept any kind of change. However, there may often be ways in designing strategic changes which will not threaten the institutional integrity of an organization.

References

Albert, S. and D.A. Whetten (1985) 'Organizational identity', *Research in Organizational Behavior,* vol. 7, pp. 263–295.

Anderson, P. and M. Tushman (1990) 'Technological discontinuities and dominant designs: A cyclical model of technological change', *Administrative Science Quarterly,* vol. 35, pp. 604–633.

Arthur, W.B. (1989) 'Competing technologies, increasing returns and lock-in by historical events', *The Economic Journal*, vol. 99, pp. 116–131.
Barney, J.B. (1986) 'Organizational culture: Can it be a source of sustained competitive advantage?' *Academy of Management Review,* vol. 11, pp. 656–665.
Barney, J.B. (1991) 'Firm resources and sustained competitive advantage', *Journal of Management*, vol. 17, pp. 99–120.
Barr, P.S., J.L. Stimpert and A.S. Huff (1992) 'Cognitive change, strategic action and organizational renewal', *Strategic Management Journal*, vol. 13, pp. 15–36.
Brunninge, O. (2005) *Organisational Self-Understanding and the Strategy Process,* JIBS Dissertation Series No. 027, Jönköping International Business School.
Burns, T. and G.M. Stalker (1961) *The Management of Innovation*. London: Tavistock.
Chreim, S. (2005) 'The continuity-change duality in narrative texts of organizational identity', *Journal of Management Studies,* vol. 42, pp. 567–593.
David, P.A. (1985) 'Clio and the economics of QWERTY', *American Economic Review Proceedings*, vol. 75, pp. 332–337.
DiMaggio, P.J. and W.W. Powell (1983) 'The iron cage revisited: Institutional isomorphism and collective rationality in organizational fields', *American Sociological Review,* vol. 48, pp. 147–160.
Eisenhardt, K.M. and J.A. Martin (2000) 'Dynamic capabilities: What are they?' *Strategic Management Journal*, vol. 21, pp. 1105–1121.
Gagliardi, P. (1986) 'The creation and change of organizational cultures: A conceptual framework organization studies', *Organization Studies*, vol. 7 (2), pp. 117–134.
Garud, R. and P. Karnøe (2001) *Path Dependence and Creation.* Mahwah, NJ: Lawrence Erlbaum.
Gebert, D. (2000) 'Zwischen Freiheit und Reglementierung: Widersprüchlichkeiten als Motor inkrementalen und transformationalen Wandels in Organisationen – eine Kritik des punctuated-equilibrium-Modells', in G. Schreyögg and P. Conrad (eds) *Organisatorischer Wandel und Transformation. Managementforschung 10.* Wiesbaden: Gabler, pp. 1–32.
Gioia, D.A., M. Schultz and K.G. Corley (2000) 'Organizational identity, image, and adaptive instability', *Academy of Management Review*, vol. 25, pp. 63–81.
Grinyer, P.H., P.D. Hayes and P. McKiernan (1988) *Sharpbenders: The Secrets of Unleashing Corporate Potential*. Oxford: Blackwell.
Hannan, M.T. and J. Freeman (1977) 'The population ecology of organizations', *American Journal of Sociology,* vol. 82, pp. 929–963.
Hatum, A.O. (2002) *Organisational Flexibility in an Emergent Economy.* Doctoral dissertation. Warwick Business School. Warwick: Warwick University.
Hendry, J. and D. Seidl (2003) 'The structure and significance of strategic episodes: Social systems theory and the routine practices of strategic change', *Journal of Management Studies*, vol. 40, pp. 173–196.
Johnson, G. (1987) *Strategic Change and the Management Process*. Oxford: Basil Blackwell.
Kimberly, J.R. and H. Bouchikhi (1995) 'The dynamics of organizational development and change: How the past shapes the present and constrains the future', *Organization Science*, vol. 6, pp. 9–18.

Leonard-Barton, D. (1992) 'Core capabilities and core rigidities: A paradox in managing new product development', *Strategic Management Journal,* vol. 13, pp. 111–125.

Lindell, P., L. Melin, H. Gahmberg, A. Hellqvist and A. Melander (1998) 'Stability and change in a strategist's thinking', in C. Eden and J.-C. Spender (eds) *Managerial and Organizational Cognition. Theory, Methods and Research*. London: Sage, pp. 76–92.

Melin, L. (1998) 'Strategisk förändring: Om dess drivkrafter och inneboende logik', in B. Czarniawska (ed.) *Organisationsteori på Svenska*. Malmö: Liber.

Melin, L. and B. Hellgren (1994) 'Patterns of strategic change processes: Two change typologies', in H. Thomas (ed.) *Building the Strategically-responsive Organization.* Chichester: Wiley.

Miller, D. and P.H. Friesen (1982) 'Structural change and performance: Quantum versus piecemeal-incremental approaches', *Academy of Management Journal*, vol. 25, pp. 867–892.

Mintzberg, H. (1978) 'Patterns in strategy formation', *Management Science,* vol. 24, pp. 934–948.

Quinn, J.B. (1978) 'Strategic change: "Logical incrementalism"', *Sloan Management Review*, vol. 20, pp. 7–21.

Reger, R.K., L.T. Gustafson, S.M. Demaire and J.V. Mullane (1994) 'Reframing the organization: Why implementing total quality management is easier said than done', *Academy of Management Review*, vol. 19, pp. 565–584.

Selznick, P. (1957) *Leadership in Administration*. New York: Harper & Row.

Schreyögg, G., J. Sydow and J. Koch (2003) 'Organisatorische Pfade – Von der Pfadabhängigkeit zur Pfadkreation?' in G. Schreyögg and J. Sydow (eds) *Strategische Prozesse und Pfade. Managementforschung 13.* Wiesbaden: Gabler.

Teece, D.J., G. Pisano and A. Shuen (1997) 'Dynamic capabilities and strategic management', *Strategic Management Journal,* vol. 17, pp. 509–533.

Tushman, M.L. and P. Anderson (1986) 'Technological discontinuities and organizational environments', *Administrative Science Quarterly*, vol. 31, pp. 439–465.

Tushman, M.L. and E. Romanelli (1985) 'Organizational evolution: A metamorphosis model of convergence and reorientation', *Research in Organizational Behavior,* vol. 7, pp. 171–222.

Wernerfelt, B. (1984) 'A resource-based view of the firm', *Strategic Management Journal,* vol. 5, pp. 171–180.

Part IV

Path-dependent Industry Development

7

Path Dependence, Modularity and the Offshoring of (Some) Physician Services

Martin Stack and Myles Gartland

7.1 Introduction

There is a large and growing literature regarding the decision making process firms undertake as they evaluate what to make versus what to buy. One group of scholars studying these issues of vertical integration uses the concept of modularity: according to Pil and Cohen (2006, p. 995) 'modularity in product design allows a firm to exploit technological opportunities through recombination, modular innovation, and outsourcing.' Modularity focuses on how complex processes can be decomposed into discrete components that can in turn be handled by the most efficient organization or transferred to the most efficient location. At the same time that new technologies and organizational processes are enabling a growing range of processes to be offshored, there is growing recognition of the regulatory and cultural barriers that often inhibit the ability of firms and entrepreneurs to fully modularize and offshore their production processes or supply chains. While many different ideas attempt to explain why specific production methods remain in place, the theory of path dependence provides a compelling framework to examine why less efficient production models may remain locked in place despite more efficient alternatives. As we argue below, path dependence is very useful in explaining regulatory and behavioural inertias that impede business practices such as offshoring and offshore-outsourcing.

This chapter integrates ideas from modularity and path dependence to explain the potential for and limitations of the offshoring of high value added physician services. While not all medical services are

amenable to offshoring, a growing number of physician activities can be viewed as modular components which, at least technically, could be performed remotely. As a result, the presumption that concerns for efficiency dictate how firms decide which activities to locate in other countries must be reconciled with complex path-dependent domestic healthcare regulatory frameworks.

The chapter begins with an overview of offshoring and outsourcing. Parts two and three examine how the ideas of path dependence and modularity can be used to explore both the opportunities in and the limitations of the offshoring of high value added physician services. Part four develops a framework that combines modularity and path dependence in assessing the firm ability to offshore medical services. The conclusion examines how these ideas can be extended to other markets.

7.2 Offshoring versus outsourcing

While the terms 'offshoring' and 'outsourcing' are used differently, there is a clear consensus that two important ideas are at work: to what extent was the work in question being performed in a firm's home country or abroad, and to what extent was a firm doing the work itself versus buying it from the market? Drawing from Stack and Downing (2005), Table 7.1 summarizes the four possible combinations of offshoring and outsourcing:

Since this chapter focuses on the offshoring of healthcare services by American heath care entities, we will use U.S. healthcare examples for each of these boxes. Box I presents the standard situation of an American hospital arranging for one of its own radiologists to read an MRI: in this scenario, the work is done in house and in the U.S. In Box II, an American hospital sets up an office in India but with its own radiologists on staff: here, the U.S. hospital continues to do the

Table 7.1 Work distribution

		Is the work done in house?	
		In house	**Outsourced**
Is the work being offshored?	Work done in the U.S.	I: In the U.S. and in-house	III: In the U.S. but outsourced
	Work done abroad	II: Abroad but in-house	IV: Abroad and outsourced

work in-house but now the work is done abroad using the new technologies that enable diagnostic radiology to be done remotely. This is an example of offshoring but not outsourcing. Box III shows a U.S. hospital that elects to contract with another hospital or a radiology group in the U.S. to read some of its MRIs: this is an example of outsourcing but not offshoring. In Box IV, a U.S. hospital that needs to have some MRIs read contracts with another hospital or a radiology group outside of the U.S.: this example combines both offshoring and outsourcing, and is referred to as offshore outsourcing.

Columns three and four address the long-standing question of when should firms 'make' versus 'buy,' that is, when should they perform a function internally versus going to the market? Rows three and four examine whether the work is being done domestically or abroad. For healthcare, as with some other services, there are other dimensions that are equally important such as whether the worker is licensed and credentialed by U.S. authorities or not: this dimension becomes more and more important as we move up the skill ladder. For example, medical transcriptions and billing agents do not require extensive licensing and credentialing in the U.S., whereas nurses and physicians do.

7.3 Path dependence and the regulation of healthcare

Paul David (1985) and Brian Arthur (1990) published several papers that serve as the foundation of the theory of path dependence: the basic assertion in these and related essays is that sub-optimal or inefficient technologies can be become locked-in as industry standards, and in instances where there are significant network effects, these inefficiencies may persist for extended periods of time (Garud et al., 2003; Stack and Gartland, 2003).

David's best-known work in this area is his discussion of the layout of the typewriter keyboard, in which he shows how the familiar top row 'QWERTY' became the industry standard despite the fact that it is an inefficient organization of keys (David, 1985). Arthur's most popular example concerns the struggle for supremacy over VCR format, and the market's choice of VHS over Beta (Arthur, 1990). Critics of path dependence have attacked the validity of these two well-known cases, and have concluded that this thesis is without merit (Leibowitz and Margolis, 1995).

The debate, of course, is not that simple: in a sense, it does not really matter whether QWERTY was or was not a good example of path dependence. The real issue is whether market processes can lock-in any

inefficient or suboptimal technologies, production processes or products. Robin Cowan has published several detailed examples of suboptimal technology path dependence. For example, in his 1990 essay, he provides an interesting example of sub-optimal lock-in, showing how light water reactors, an inferior technology, triumphed over superior alternatives.

Choi and Stack (2005) present another extension of path dependence, arguing that consumer behaviours can become locked in to a suboptimal industry standard 'once a product has become established as an industry standard, and once consumers are locked into a behavioural routine, they will be less likely to try a rival product even if, over time, it proves superior' (Choi and Stack, 2005, p. 80). They integrate the related ideas of path dependence, lock-in and switching costs to help explain how – historically and strategically – a particular style of beer came to dominate the American market. They called this process an example of behavioural lock-in, to distinguish it from the cases of technological lock-in adduced by David, Arthur, and Cowan.

Behavioural lock-in is particularly important in this chapter, as one dimension of the lock-in concerns physician interactions. Section 7.2 used diagnostic radiology as an example in discussing examples of offshoring and outsourcing: while diagnostic radiologists do not interact directly with most of their patients, they do develop deep formal and informal relationships with the primary care and emergency room physicians with whom they work. Any efforts to offshore, outsource, or offshore-outsource diagnostic radiological services must take into account the path dependent behavioural dimensions of these physician-to-physician relationships since they have proven very difficult to alter. As with professionals in many fields, primary care physicians (PCPs) often develop levels of comfort in and confidence with radiologists who are known quantities: absent compelling reasons, it will often be difficult to persuade PCPs to work with new radiologists with whom they have even fewer personal and professional connections. This is true even for intra-national changes, so efforts to offshore this work internationally will be even more difficult. While professional confidence can be established over time, the very notions of path dependence and behavioural lock-in highlight that it is often difficult to break free of established formal and informal relationships.

The common link across these path dependence stories is how inefficiencies can become locked into place. This insight can also be extended to explain why outdated government rules and regulations remain in place despite radical changes in the environment for which they were

designed. Political scientists commonly study how institutions encourage individuals to act in ways that lock-in a particular path of policy development creating societal commitments that may be quite difficult to reverse (Pierson, 2000). Regulatory lock-in, then, focuses on how well intentioned governmental policies can become inappropriate as the initial conditions change. In healthcare, a powerful example of regulatory lock-in concerns the requirements regarding the licensing and credentialing of physicians.

Regulation impacts all service industries to some degree, but healthcare has some of the stiffest rules of any market. According to Kovner and Salsberg (2002), the U.S. 'has relied heavily on licensure, accreditation, and certification of health professionals to assure that healthcare was high quality ... Most health professions are licensed by states. Licensure requires graduation from an approved school, passage of an examination, and, in certain cases, a minimum period of practical experience' (Kovner and Salsberg, 2002, p. 95). Referring back to Figure 7.1, this regulatory arrangement basically holds a box I view of the delivery of physician services. Yet, over the course of the twentieth century, a number of technological advances have emerged which make boxes II, III and IV equally feasible: the problem is that these alternative modes of care delivery run up against a healthcare regulatory system based on a box one model of how physician services should be delivered. According to Barnes et al. (2004, p. 373), 'once a particular behavior is embedded in organizations (for whatever reason), a strong status quo inertia may discourage other behavior. It takes tremendous effort to change, and individuals may simply be uninterested (even when they suspect it is necessary).'

Regulatory lock-in challenges other health systems as well. The European Union is dealing with an intra-border regulatory struggle that is similar in many ways to the states licensure issue in the U.S.: as in the U.S., this struggle has impeded the cross border delivery of many 'transportable' healthcare services.

One of the biggest challenges to this traditional model has been the advent of telemedicine, which is defined as 'connecting geographically separate healthcare facilities via telecommunications, video and information systems' (Klein and Manning, 1995, p. 35). The problem is that while the states have the regulatory power to license who practices medicine within its boarders, telemedicine is presenting new opportunities for practitioners to treat patients remotely, that is, across state lines: as Klein and Manning (1995, p. 39) put it, 'The technology is moving nimbly to satisfy the demand for telecommunications and

video that connect geographically separated healthcare organizations. Unfortunately, the legal and regulatory environment has not progressed as rapidly.' They highlight a series of path-dependent legal impediments such as state licensure laws that are slowing down the expansion of telehealth. These impediments exist even for hospitals trying to move from Box I to Box III; the barriers to introducing Boxes II and IV strategies which involve offshoring are much more substantial.

For specialists such as radiologists, the regulatory requirements are even more extensive: not only must they be licensed in the states in which they seek to practice, they have to meet specific certification and accreditation criteria required by their specialty association, in this case the American Board of Radiology (www.theabr.org). Together, these licensure, certification, and accreditation requirements create significant regulatory path dependencies that inhibit the offshoring of radiology (White, 2002). While designed historically to establish baseline competencies for U.S. physicians, in a rapidly changing world in which some healthcare services are capable of being delivered internationally, they act-intentionally or not-as a restraint on trade.

7.4 Modularity and physician services

To explain the forces that are simultaneously facilitating and resisting the offshoring of physician services, we use the complementary concepts of path dependence and modularity. Langlois (2000) refers to modularity as 'a very general set of principles for managing complexity. By breaking up a complex system into discrete pieces – which can then communicate with one another – one can eliminate what would otherwise be an unmanageable spaghetti tangle of systemic interconnections' (Langlois, 2000, p. 5). The very process of modularization helps firms address the age old problem of what to produce versus what to acquire from the market. As technologies and organizations evolve, the notion of what a firm should build versus buy necessarily changes as well. Increased modularity enables firms to specialize in specific areas, thereby enhancing their economies of scale and scope. This also creates a positive feedback loop as specialized firms are rewarded for innovation as long as they maintain a common interface. Advocates cite the modular home stereo system as an example (Langlois, 2000). As entrepreneurs introduce new equipment (e.g., a high definition DVD player), stereo owners can replace an older part (e.g., a VCR) without having to change all the other components of their system. This suggests that micro product level innovation (a single component) is

easier than innovation of an entire system which tends to suffer from much stronger degrees of path dependence. Some healthcare services manifest modular characteristics since distinct components can be innovated without completely disrupting the entire process. As the module interfaces become increasingly well defined and standardized, innovation continues and the transactions costs of buying from the market begin to fall as well. Thus, advocates of modularity highlight its technical and organizational efficiencies (Langlois and Robertson, 1995; Garud et al., 2003).

Advances in technology have greatly increased the range of physician services that can be modularized and, as a result, outsourced, offshored, and offshore-outsourced. The list of that risk specialties includes psychiatry, pathology, cardiology, dermatology, ophthalmology, diagnostic radiology and, more generally, any specialty which utilizes digital imaging and digital monitoring for diagnoses and interpretation. The key dimension is the degree to which the healthcare professional can be separated from the delivery of the service: the greater the degree of separability, the more amenable it is to modularization. Yet, before we explore the challenges confronting new delivery mechanisms for physician services, we need to review exactly how some of these activities can be modularized, as it is this feature that makes offshoring and offshore-outsourcing possible.

Let us briefly consider the example of radiology. Over the course of the past century, technical advances have steadily expanded the scope of radiology, which is currently divided into two sub-disciplines, diagnostic radiology, and interventional radiology. Diagnostic radiologists use a wide and sophisticated array of technologies to diagnose a series of medical conditions; interventional radiologists use radiation and other processes to help treat their patients. One fundamental distinction between them is the degree to which their work requires interaction and contact with patients. Interventional radiologists interact directly with patients and are primary care and/or specialty physicians: proximity is a necessity. In contrast, diagnostic radiologists utilize a number of technologies to help other physicians better understand the nature and extent of specific illnesses and conditions: their work requires them to interpret digital images that are transmitted to them by hospitals, clinics, and physicians. Typically, most diagnostic radiologists develop strong personal relationships with primary care and specialty physicians who order these scans and images, but they do not usually interact directly with the patients, and their interactions with the physicians who order radiological scans are most commonly by email or phone.

As a result, there are several modular dimensions to this work. First, with the introduction of digital imaging and the widespread dissemination of broadband technology that is necessary to transmit these images (from the office in which the procedure is taken to the diagnostic radiologist's office), proximity to the patient, the hospital, and to the supervising physicians is no longer necessary. These technical developments have, in essence, enabled a series of once tightly connected physician services to be sub-divided into manageable, discrete pieces; within certain licensure constraints. Thus, diagnostic radiology can be offshored from and to any countries with the necessary technology and supporting communication infrastructure.

In addition, in the United States there is a second factor that further facilitates physician modularization: a highly decentralized healthcare system. Unlike other developed countries, most U.S. physicians work collaboratively with rather than directly for the complex web of healthcare providers and payors which includes the federal and state governments, the array of public and private hospitals, and the growing number of managed care companies. While diagnostic radiologists work with all of these entities, most of them are actually employed by separate radiology practices. This separation is true for most primary and specialty physicians in the U.S., and this unique contracting relationship – somewhere between market and in-house – has over time created a degree of modularity for U.S. physician services that is rather different from other developed countries.

Together, the combination of new digital and communication technologies, and a physician system that has kept doctors somewhat separate from the healthcare entities with which they interact on a daily basis, have created a modular system that is ripe for outsourcing and offshoring. Thus, modularization is a process that diminishes path dependence and stimulates innovation. Yet, change is often threatening, and technical developments which favor modularization may run into a number of barriers which inhibit the creation of these distinct modular components. As a result, Garud et al. (2003, p. 7) emphasize that while 'modularity is a rich entry point to a broader set of issues cutting across technological, organizational, and strategic domains' they also firmly recognize that 'we cannot talk about the benefits of modularity without acknowledging the socio-political process involved in the shaping of industry-wide standards.' This last point highlights how pressures to develop modular systems often run into path dependent regulatory forces.

7.5 A framework for integrating path dependence, modularity, and offshoring

Thus far, we have reviewed how the theories of modularity and path dependence explain the competing factors that are simultaneously facilitating and impeding the offshoring of physician services such as radiology. The next step is to develop a framework that integrates path dependence, modularity, and offshoring. While the model may be applied to a range of distinct markets, we apply it to the specific case of physician service offshoring.

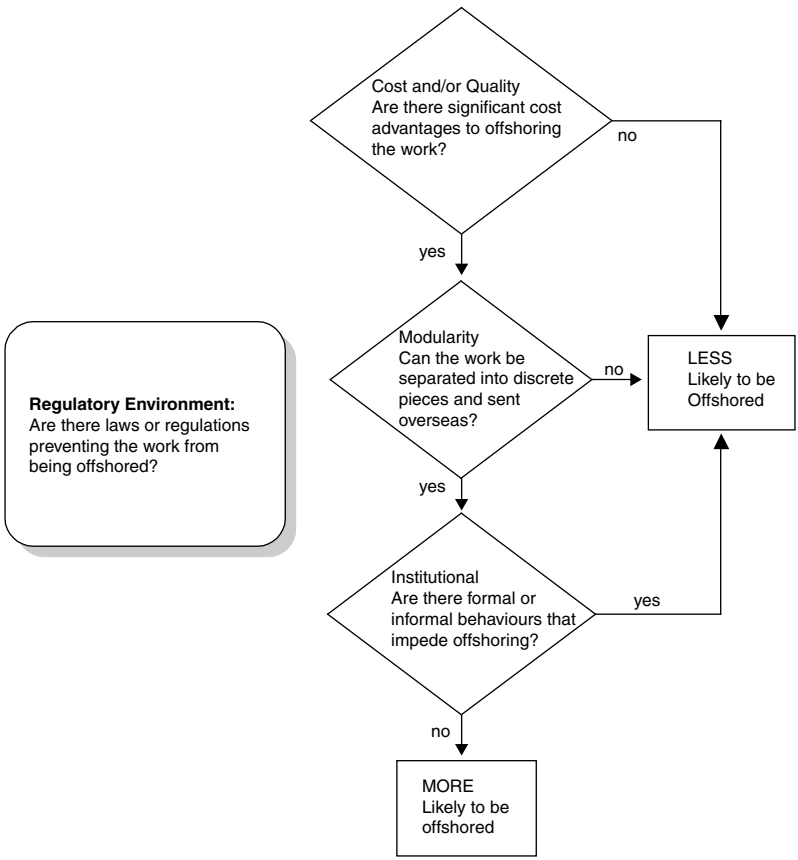


Figure 7.1 Framework for integrating path dependence, modularity, and offshoring

The story typically begins in the top box with motivation: can offshoring or offshore-offshoring provide significant opportunities for lower costs and/or higher quality? It is almost always easier to use local or national resources, so any explorations of going abroad for the delivery of a good or service (what is often referred to as 'the globalization of production') must justify why foreign producers may be preferable.

Let us first consider cost. At the aggregate level, spending on healthcare in most nations has been rising rapidly for decades. As a percentage of GDP, health spending in the U.S. has risen from 7 per cent in 1970 to 16 per cent in 2007, when national healthcare expenditures totalled over $2.2 trillion dollars. Interestingly, the range of services most amenable to offshoring are higher paying specialty services such as radiology, pathology, cardiology, and dermatology. Consider the case of diagnostic radiology. The American Medical Group Association reported recently that the average diagnostic radiologist earned nearly $365,000 (Radiological Society of North America, 2005). In an era of escalating healthcare costs, rising salaries for one of the highest paid physician specialties makes offshoring of radiology to countries with lower wage structures an attractive proposition: one healthcare firm in Bangalore India pays its radiologists who work for U.S. clients $30,000–$100,000, well above the average Indian salary for radiologists but still low enough for the firm to charge about half what radiology groups in the U.S. charge (Pollack, 2003). Clearly, healthcare costs such as wages provide a significant incentive for exploring offshoring opportunities.

Yet, firms may go overseas even if costs are the same or higher: there may be key resources that are more widely available in other countries, or that can be located in other countries to take advantage of time differentials. By working with radiologists in Europe and Asia, primary care physicians in the U.S. could satisfy potential quality concerns by utilizing less overworked staff during their daytime hours. Depending on where these overseas radiologists are located, their salaries may be lower (if, e.g., they are located in India), or higher (if, e.g, they are located in Switzerland or Australia) than prevailing salaries in the U.S. Figure 7.1, however, shows that while cost or quality differentials are necessary, they are far from sufficient. Thus, moving on Box II, the next step is to detail whether the work can be offshored. That is, are the services modular in the sense that a healthcare firm could, in theory at least, separate out discrete steps in the delivery of care? Again, radiology provides an illustrative example. As discussed in Section 7.3, recent advances in medical technology have further sundered the need

for diagnostic radiologists to be physically near the primary care and other specialty physicians with whom they consult. Today, diagnostic radiologists, along with a growing number of other physician specialists, depend more on reliable communication networks than on shared physical facilities. This is not to suggest that all physician services have this modular dimension; many do not. In fact, most paediatric, obstetric, primary care, and emergency room care cannot be provided remotely and thus are not at risk for offshoring.

But, for those services that are modular and which have significant cost differentials, then Figure 7.1 directs us to box III, and asks if there are path dependent formal or informal institutional behaviours that may impede modularizing and offshoring key physician services. This moves the discussion from what is economically and technically feasible to what is actually possible in behavioural terms. Many PCPs prefer to work only with those radiologists with whom they have a personal or professional connection. While radiologists and other physician specialists who do not need to interact in person with consulting physicians can, in theory, locate in many places, they still tend to base their offices in the same medical complexes. This is an excellent example of the power of informal connections – in field after field, many professionals continue to base their operations near other professionals. Box III, then, represents those behavioural path-dependent forces – formal and informal – that may impede the ability of firms to modularize some of their constituent services.

Finally, Figure 7.1 shows that the discussions of Boxes I–III take place within a broad regulatory environment. So, a firm may find that it is economically and technically feasible to offshore some of their physician services, and they may be able to surmount the formal and informal institutional and behavioural barriers as well; but, all business and economic activities take place within broader path-dependent and socio-politically determined regulatory environments. For example, in the United States, the Federal government is a major payer of healthcare services through the Medicare and Medicaid programs. Under these programs, the government essentially does not reimburse providers for services rendered outside of the United States. So, a hospital that finds it economically, technically, and institutionally feasible to contract for offshore telemedical modular services may ultimately not be able to do so if it will not be compensated for these services by one if its largest reimbursers. As noted in Section 7.3, there are strong path dependencies in the evolution and functioning of industry-specific regulatory frameworks.

The American College of Radiology (ACR), along with a number of other professional associations, has weighed in on whether facilities in the U.S. should be allowed to contract with radiologists based abroad. There are many compelling reasons for regulators and healthcare professionals to review and often challenge efforts to modularize and offshore amenable physician and nurse services. Perhaps the most important concern involves quality: how can we vouch safe the quality of physician services rendered not only remotely but outside of the United States? Current regulations in the United States allow facilities to utilize offshore teleradiology procedures as long as the radiologists are licensed by the ACR. Drawing on the path dependence and modularity ideas of this chapter, several teleradiology firms are currently trying to balance the cost, quality, technology, behavioural, and regulatory factors that collectively serve to determine the viability of offshoring and offshore-outsourcing healthcare services. Two firms currently fit the model of Figure 7.1 very well, while a third firm is at present trying to surmount some of the path-dependent regulatory requirements. Teleradiology Solutions (www.teleradiology.com) illustrates how a company can reconcile telemedicine's modular opportunities with its path dependent regulatory and behavioural constraints. This firm has a staff of American Board of Radiology certified radiologists who are currently licensed in over 20 states. All of these radiologists have returned to India after practicing for varying amounts of time in the U.S. Teleradiology Solutions has further enhanced its competitive position by becoming the first Indian company to become fully accredited by the Joint Commission International (JCI), the independent healthcare organization in the United States which evaluates and accredits healthcare entities and program (Syed, 2005). Expenses in India are a fraction of what they would be in the United States, so Teleradiology is able to compete very aggressively on the cost dimension. As all of the radiologists are ACR certified, the company is able to satisfy, to varying degrees, the formal and informal regulatory and behavioural constraints that impact physician offshoring initiatives. Yet, while this has so far proven to be a successful business model, it is clearly a small niche – there are only a small number of U.S. board certified radiologists who are willing to relocate to lower cost countries such as India.

Another successful approach comes from the firm Nighthawk Radiology Services (www.nighthawkrad.net) which hires in U.S. licensed and credentialed radiologists and locates them in Zurich, Switzerland, and Sydney, Australia. Clearly, this example of offshoring is not about saving costs, as the company is hiring in certified personnel and locating

them in strategically selected time zones so that the work will typically be performed by radiologists working during their day hours. In addition to its Box I emphasis on quality, it believes that offshoring modular physician services faces behavioural path dependent constraints: that is, it reasons that American PCPs will feel more comfortable offshoring services such as radiology to other western countries than to developing countries such as India. That is, this approach believes that Box III behavioural issues regarding U.S. physician comfort levels regarding the countries they are interacting with using telemedicine will typically reflect Box I concerns about quality not cost savings. At this stage, both Teleradiology Solutions and Nighthawk Radiology are doing well, so it will be interesting to see if either model proves more successful in the long run.

A third approach comes from Wipro, one of India's best known high technology companies which also has a healthcare division based in Bangalore.[1] Rather than restrict itself to the rather limited supply of Indian radiologists who are licensed and credentialed in the United States (the Teleradiology Solutions approach), Wipro has chosen to hire in highly trained Indian radiologists, some of whom may have studied or worked abroad in the U.S. or the U.K. Yet, while this may increase the supply of radiologists from which to draw, it currently circumscribes the level of work these radiologists can perform for their U.S. clients. If the Wipro radiologists are not licensed in a U.S. state and credentialed by a U.S. hospital, they are not allowed to perform full radiological reads; instead, they provide a range of complementary services that can speed the work flow of radiologists in the United States. A few examples include providing 3D anatomical models for MRIs and CT angiographies, and providing preliminary radiological reads that can greatly increase the efficiencies of U.S. based radiologists. In stark contrast to Nighthawk Radiology, the Wipro approach weighs Box I much more heavily than Box III and the surrounding regulatory environment: or, to put it another way, it believes that cost issues will gain in importance over time, and that they may overcome formal and informal behavioural path dependencies. If efforts to create a more uniform global trading mechanism for health services through the General Agreement on Trade in Services take hold, the approach by Wipro may ultimately be rewarded.

As these examples illustrate, services rendered in the forms of Boxes I and II do not necessarily need to be modularized, and thus they suffer from the least amount of path dependent resistance. Box III requires modularization, but it tends to face low degrees of regulatory lock-in as

it is being done within the borders of the same country. Box IV requires the greatest amount of modularity and it generates the highest levels of path dependent resistance: all three of the firms examined here fall into Box IV, though the regulatory and behavioural path dependencies each faces vary according to their specific business model.

7.6 Future directions

Technology induced modularity has greatly increased the ability of diagnostic radiologists to operate in other countries; yet, the ability of firms to modularize these and other professional healthcare services to offshore locations have been impeded by informal behavioural path dependencies that influence who domestically-based healthcare professionals will work with and by formal regulatory path dependencies which specify the licensing, credentialing, and legal requirements which healthcare offshoring must satisfy. This chapter highlights the factors that are both facilitating and restraining the offshoring of high value added physician services. It argues that this discussion can be readily extended to a large and growing range of healthcare services. For example, many healthcare plans in the United States have begun phone and email-based 'Ask a Nurse' programs as a way to provide lower cost yet high quality medical consultation programs for their members. As these programs use on-line databases and phone and email based interactions, these services have already been modularized: so, the next question is how path-dependent providers, patients, and regulators are. Will the providers and patients adjust their own behaviours to allow for these nurse programs to take advantage of potential cost and quality improvements by locating overseas? Will the regulators modify existing requirements regarding licensing and legal hurdles if quality levels can be vouchsafed?

While the focus of this chapter has been on healthcare, it can clearly be extended to a number of other service sector parts of the economy, especially those which are characterized by the dual tension of increasing technical modularity and path-dependent regulatory frameworks. Alan Blinder (2006) recently argued that

> The fraction of service jobs in the United States and other rich countries that can potentially be moved offshore is certain to rise as technology improves and as countries such as China and India continue to modernize, prosper, and educate their work forces. Eventually, the number of service-sector jobs that will be vulnerable to competition

> from abroad will likely exceed the total number of manufacturing jobs. Thus, coping with foreign competition, currently a concern for only a minority of workers in rich countries, will become a major concern for many more. (Blinder, 2006, p. 120)

Continued improvements in communication and computer technologies will steadily expand the range of duties that can be modularized. Yet, the simple projection that Blinder makes does not adequately take into account the complex array of behavioural and regulatory path dependencies that such efforts will encounter. The real question is how firms and countries, consumers and professional associations handle the increasing tensions of modularizing and path dependence as they relate to offshoring and offshore-outsourcing.

Note

1. This information is based on interviews conducted at Wipro in Bangalore in June, 2005.

References

Arthur, B. (1990) 'Positive feedbacks in the economy', *Scientific American*, vol. 262, pp. 92–99.

Barnes, W., M. Gartland and M. Stack (2004) 'Old habits die hard: Path dependency and behavioural lock-in', *Journal of Economic Issues*, vol. 38 (2), pp. 371–377.

Blinder, A. (2006) 'Offshoring: The next industrial revolution', *Foreign Affairs,* vol. 85 (2), pp. 113–128.

Choi, D. and M. Stack (2005) 'The all-American beer: A case of inferior standard (taste) prevailing', *Business Horizons*, vol. 48, pp. 79–86.

Cowan, R. (1990) 'Nuclear power reactors: A study in technological lock-in', *Journal of Economic History*, vol. 50, pp. 541–567.

David, P. (1985) 'Clio and the economics of QWERTY', *American Economic Review, Papers and Proceedings*, vol. 75, pp. 332–337.

Garud, R., A. Kumaraswamy and R. Langlois (2003) *Managing in the Modular Age: Architectures, Networks and Organization*. Malden, MA: Blackwell.

Klein, S. and W. Manning (1995) 'Telemedicine and the law', *Healthcare Information Management*, vol. 9 (3), pp. 35–40.

Kovner, C. and E. Salsberg (2002) 'The healthcare workforce', in C. Kovner and S. Jonas, *Healthcare Delivery in the United States*, 7th edn. New York: Springer Publishing.

Langlois, R. and P. Robertson (1995) *Firms, Markets and Economic Change*. New York: Routledge.

Langlois, R. (2000) *Firms, Markets and Economic Change*. University of Connecticut Working Paper.

Leibowitz, S. and S. Margolis (1995) '"Policy and path dependence" from QWERTY to Windows 95', *Regulation*, vol. 18, pp. 33–41.
Pierson, P. (2000) 'Increasing returns, path dependence, and the study of politics', *American Political Science Review*, vol. 94 (2), pp. 251–267.
Pil, F. and S. Cohen (2006) 'Modularity: Implications for imitation, innovation, and sustained advantage', *Academy of Management Review*, vol. 31 (4), pp. 995–1011.
Pollack, A. (2003) 'Who's reading your x-ray?' *New York Times*, November 13.
Radiological Society of North America (2005) 'Salaries flat for interventional diagnostic radiologists', from http://www.rsna.org/Publications/rsnanews/oct05/salaries.cfm.
Stack, M. and M. Gartland (2003) 'Path creation, path dependency and alternative theories of the firm', *Journal of Economic Issues*, vol. 37, pp. 487–494.
Stack, M. and R. Downing (2005) 'Another look at offshoring: Which jobs are at risk and why', *Business Horizons*, vol. 48 (6), pp. 513–523.
Syed, F. (2005) 'Teleradiology solutions becomes 1st JCAHO accredited Indian body', *Express Healthcare Management*, July 16–31.
White, P. (2002) 'Legal issues in teleradiology – distant thoughts!' *British Journal of Radiology*, vol. 75, pp. 201–206.

8

The Horse's Hoof: Social Identity, Path Dependence and Path Creation in the Equine Industry

Kenneth W. Koput, Michael W. Conaway and David Olson*

8.1 The equine industry – change, but predominance persists

The horse industry in America has undergone a sea change over the latter part of the twentieth century. Whereas horses were once either tools of the rancher's trade or the stable of the affluent, recent surveys indicate that the vast majority of horses owned in America today are used for recreation and pleasure by persons from all walks of life and socio-economic status (American Horse Council, 2005). Racing and working horses now account for less than 20 per cent of all horses in the U.S. Although the new ways in which horses are kept and used alter the functionality of shoes, 95 per cent of horses are still shod with the all-purpose steel 'keg' shoe that has been the dominant design since the 1800s (*Rubber & Plastics News*, 2005).

As horses are kept more for recreation and showing (and often simply as pets), than for work or racing, individual owners have come to rely on an array of support specialists. A new industry has emerged, made up of breeders and breed registries, trainers, farriers, equine veterinarians, boarding stable owners/managers, feed and tack suppliers, nutritional specialists, therapists, and clinicians supported by a variety of associations, magazines, websites, and a cable TV channel.

Table 8.1 contains percentages of respondents by categories of reasons for owning horses from a survey of local horse owners in the American Southwest, where the image and identity of a horseman has a strong tradition.[1] The image of the lone rancher responsible for the breeding, training, and upkeep of his own horses no longer describes the typical

horse owner in Arizona. As seen in Table 8.1, only 10 per cent of survey respondents keep horses for work, and another 6 per cent for other business or investment reasons.

Table 8.2 shows the demography for the owners in our survey. Again, the shift is apparent, with the average owner being a college-educated, married woman of 50 years in age. Occupations vary widely, but the largest group hold white-collar or intellectual jobs.

Although the picture of the working rancher is fading, tradition lingers, as the tack and techniques of the western horseman remain key tools of the trade and the source of nostalgia (Ensminger, 1991). Let's say a few things about that horseman, then, as the identity is clearly a starting point in the psyche of our population. As an icon of the American West, the image of the lone rancher calls forth the notion of independence, individualism, and righteousness. These are all traits of the western horseman and all are symbolized by the horse (Ensminger, 1990).

Herein lies a contradiction in the identity of new owners that has fuelled growth in an otherwise stagnant equine industry: their own

Table 8.1 Why own horses

Reasons for owning horses	**Per cent respondents**
work	10
hobby/recreation	75
business/investment	6
competing	19
lifestyle	27
companion/love horses	70

Table 8.2 Who owns horses

Demographic	**Means/percentages**
age	50
gender	82% women
marital status	70% married
education	16.5 years
occupation percentages	
horse-related	12
white collar/intellect	40
blue collar/labor	20
social/community service	20
student/retired	23

self-image as western horsemen/women is perpendicular to their lack of knowledge and their reliance on others. The new horse owner is, often, in search of a social identity as a horseman. The dynamics of that search determine path breaking and path creation in the equine industry. As we will describe throughout this chapter, social identity shapes the adoption and dissemination of new technologies and practices in the field. Social movements and social networks shape how information gets shared and used, helping to establish possible identities.

Overall, the horse industry in the U.S. has a direct economic impact of $39 billion dollars and 1.4 million full-time jobs, with another $60 million in indirect economic contribution to the nation's GDP (American Horse Council, 2005). The horseshoe industry is itself a big business, with an annual sales base of $80 million for shoes alone and upwards of $2 billion per year when farrier services and supplies are included (*Rubber & Plastics News*, 2005; *Fortune*, 1993). The shoeing industry is undergoing variation as a result of the demographic shift. Farriers, once numbered in the hundreds, have grown in number to approximately 30,000 in the U.S. *Fortune* magazine reported in 1993 that start-up companies were adding a new look: the decade saw hundreds of new firms, 600 different kinds of shoes, and 50 different kinds of nails introduced (*Fortune*, 1993).

There are a number of reasons that the vast majority of horseshoes sold in the U.S. are, nonetheless, the same round-toe, iron design that has shod horses for centuries. Paramount is the fact that this is a path-dependent technology and practice, as we'll show in Section 8.2. Economic factors point to type-3 path dependence, as low economic switching costs and lack of installed base prohibit a full explanation for the persistence of a sub-optimal design (Liebowitz and Margolis, 1990, 1995; Gartland, 2005). Once a lock-in occurs, interests align with the predominant technology or practice, such that multiple processes persist, and can operate in parallel, supporting the technology or practice even if the original economic factors providing positive feedback change (Mahoney, 2000). In the horseshoe, we find a social trap, in which institutionalization, power, inertia, and social identity keep horse owners from breaking the old path (David, 1985).

Again, our survey provides some insight. Only 13 per cent of respondents say that cost is a factor in determining how to shoe their horses (see Table 8.3). The performance of the shoe fares as more important, but still is not cited by a majority. The welfare of the horse is the most cited factor, followed by the use of the horse. The ability of the farrier is a factor for a majority of owners in our sample. The owners' own

Table 8.3 What owners consider in shoeing horses

Factor in how shod	Per cent respondents
cost	13
performance	32
use of horse	65
welfare of horse	76
own beliefs	10
what others do	5
farrier's ability	57

beliefs and what others do are also mentioned. Multiple processes are at work.

There is a strong tradition in the look of a horse's foot. The appeal of this look is deeply ingrained. Consider the logo for 'The Horse's Hoof'– an organization that promotes keeping horses unshod (www.thehorseshoof.com). It shows the hoof not as the natural, somewhat ugly one found on free-roaming horses, but as the idealized, long-toed standard that is readily obtained only with round-toed shoes. The point here is that this standard has become so taken-for-granted that even a barefoot site has embraced it, without a second thought.

The long-toed standard promotes unnatural movement in the horse, providing job security for trainers and a basis for associations to establish control over competition and judging. For the associations, the control over competition captures revenue from registrations, dues, and premiums. Even trainers that otherwise practice natural horsemanship are prone to taking the long toe for granted. When the unnatural nature of the long toe is pointed out to them, they are visibly taken aback.[2]

Shoeing itself is still typically functional, even if the long toe is not. Once, shoes were needed to protect the hoof from overuse and wear. Today, the hoof needs protection from underuse. The rub here is that the new owners, while looking to connect to nature or tradition through their horses, find their horses in unnatural conditions and nontraditional uses (McBane, 1992). Given domestic conditions and light duty, the hoof becomes soft and the coffin bone sinks, leading to lameness. The old design, in promoting a long toe, does not allow the toe to break-over nor does it allow for sole pressure. In cases where the horse gets a lot of work in frequent training and competition, the round-toed shoe may be of some utility. More often, the long toe and lack of sole callous are a leading cause of navicular syndrome, and providing

Table 8.4 How owners use horses

How horse is used	Percentage of horses
work	9
competing	21
breeding	13
showing	11
trail riding	59
not ridden	20

toe break-over and pressure to the digital cushion are important in the treatment of laminitis (Bowker et al., 1993; 1998).

Table 8.4 shows how owners in our survey use their horses. Just 9 per cent use them for daily work. Some fraction of those used for competition, showing, or trail might see use nearly every day, but most in those categories are likely used on a weekly or biweekly basis, or only sporadically. The remainder of the horses are not ridden at all.

We next retrace the historical development of the horseshoe to see that the defining characteristics of path dependence indeed fit. Then we will take up the question of whether this path dependence can be overcome and new paths created.

8.2 Path dependence

The all-purpose steel shoe has the key elements of path dependence (Arthur, 1989). There are three to review: i. early variation, ii. historically distant standardization, and iii. positive feedback with externalities.

8.2.1 Early variation consisting of functional adaptations to idiosyncratic problems

The ancestry of horses has been traced to equids that roamed the Mongolian Steppes. The first record of the domesticated horse appears in Egypt around 1500 BC, during the time of Seti I. Other early records indicate the use of horses in the armies of Alexander the Great (350 BC), Hannibal, and Mithridates (100 BC). It seems apparent that any early use of horse as peaceful transport was quickly displaced by the need for horses in battle, where the advantages were proven by the fortunes of the armies that had them versus those that did not (Evans, 1990).

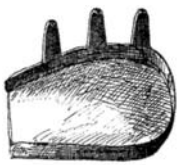
Before shoeing, horses in battle had their feet run sore and armies were delayed while the horses were idled so their hoofs could grow (Heyerming, 2008). A rawhide boot was used as a treatment for worn-out

hoofs in hopes of getting the horses back into battle more quickly. The rawhide was replaced, around the beginning of the Common Era, by a light metal sandal – such as the Roman 'cure-shoe' – attached by ropes through loops (see Figure 8.1; for images of other shoes described, see Zippelius, 1891). At the same time in history, other designs proliferated: non-Roman shoes included those of Southern Europe, marked by a thin plate with staved rim, and the Mongol 'calk shoe'.

It was not the case that an initial design was introduced and then diffused and locked-in by virtue of either the hazard of change (inertia) or the cognitive process of taken-for-grantedness (institutionalization).

Roman cure shoe

Hunnish shoe, circa 500 CE



Shoes circa 1000 CE

Asiatic (L), German, and Turkish (R)

British, circa 1700 CE

Current design, dominant since 1800s

Figure 8.1 From early variation to standardization

Source: Courtesy of Project Gutenberg's Scientific American Supplement No. 819, used freely under the terms of the Project Gutenberg License (www.gutenberg.net).

The variation in early designs applies also to functional adaptations that occurred along the way. As an example, we can consider the problem of toe protection that arises when shoes began to be used during movement rather than recuperation; this occurred around 400 CE. In movement, the toe is thought to propel the horse – hence, toe protection is a key for a shoe purported to be used in and for battle. All the therapeutic 'shoes' lacked a secure toe and produced pastern wear. Forms intended to redress this problem include the Norman 'gripe shoe,' with a toe cap nailed on, and an Asiatic shoe, circa 1000 CE, made from tin with clips that gripped into the side of the hoof wall. Yet, earlier designs foreshadowed more modern variations, but they did not lock- in. A Hunnish example from 500 CE was made of copper or tin, with punched holes that allowed sunken nail heads – preventing the nail heads from wearing off and the shoes from coming loose. This was 500 years before the Norman and Asiatic shoes just described.

The lack of diffusion and common adoption of what today appear to be superior early variations would not be surprising if they developed in isolation or secret. Yet, this is not the case, since the migration of nations occured exactly during this period 400–1000 CE, between the emergence of the Hunnish shoe and the Asiatic shoe. A German shoe circa 1000 CE gives us some evidence that these variations did diffuse and yielded reciprocal influence, as it has the Hunnish influence rather than being like shoes from Southern Europe. Early variations were not isolated, but were adapted rather than adopted as found.

Immediately after the migration of nations, the Art of Shoeing came into practice (Clark, 1831). Previously, the individual horse rider did shod his animals. At some point around 1000 CE, armies assigned the task of shoeing to specialists and as a result of the mixing during the migration of nations, this practice was widely adopted. The impact of this professionalization is noticeable, as iron becomes more and more common, as do shoes fastened by nails, sunken heads, and the placement of nails to protect the membranes of the hoof. Even so, it would be erroneous to extrapolate that shoeing had become standardized (Heyerming, 2008). If we look at the period of the Crusades (Severin, 1989) 1100–1300 CE, we can distinguish shoes of varying designs: Spanish, English, Southern European, and still others from South Russia and North Africa (Zeppelius, 1891).

8.2.2 Historically distant standardization

We now step back, as a key event occured when the Catalan Forge was introduced in 800 CE. Spanish-Gothic and Turkish-Greek shoes are

known examples of early cast shoes. We emphasize the point that even while the forge and cast shoes came along centuries earlier, their use in shoeing horses was limited. It was not until 1000 CE that the Hunnish design was cast, in a French shoe. An Asiatic shoe of the same period was also cast, but the influence of the earlier plate-shoe made the Asiatic shoe quite different at this time.

We finally see some standardization after the crusades, as key points of different designs come together in shoes from Southern Germany, circa 1300 CE. These are a plate-like shoe of southern character, but with an open-heel, broad toe, rear calks, and Hunnish holes for sunken nail heads. This design lasted until 1500 CE as the move towards a standard was reinforced by the diffusion of the Art of Mechanics and blacksmithing during the Renaissance. Farriery was introduced as an official profession around 1300 CE, influenced by the Romans and Normans (Fleming, 1869). And the manufacturing of shoes became widespread as the Catalan Forge, in place since 800 CE, gave way to the invention and – more importantly – improvement of the blast furnace (first invented 1000 CE, improved 1400 CE). By 1500 CE, then, we have historical events – big and small – creating conditions for positive feedback with externalities.

8.2.3 Positive feedback with externalities

Positive feedback is about how use leads to more use and a particular design gains a foothold or installed base. The economics of positive feedback come in self-reinforcing mechanisms (Arthur, 1989). Positive feedback is given substance in our story by:

- large fixed costs in manufacturing plant and equipment;
- learning effects for farriers in understanding the anatomy of hoof and being able to see inside of foot, by doing rather than by observing;
- coordination effects in the professionalization of farriery and in military campaigns; and
- adaptive expectations on the part of owners. Once shod, it becomes hard to go unshod as the natural toe and callous take time to re-develop.

Externalities also provide positive feedback. Externalities are really about how my adoption influences yours, how third parties support adoption and why a standard develops even though there are important regional factors that might favor different designs. The substance of

externalities in our story derives from:

i. Competitive adoption of armies and other military arrangements including the earliest ancestor of the motor pool. The specialization and formalization of the shoe as a position provides technical interrelatedness.
ii. Widespread manufacturing provided by the blast furnace, then the mass production machine, facilitated by transportation networks that allowed widespread distribution, provide economies of scale.
iii. These elements combined with a mindset of trainers and farriers in terms of what a hoof should look like, and this was driven by the professionalization and registration of farriers. The historical use of the horse in battle favored a long toe, which would have been thought to propel the horse's movement. Once this mindset becomes ingrained, it acts as an irreversible investment.

Picking up our history in the 1600s, positive feedback took hold as shoes throughout Europe began to take more features of earlier designs including Hunnish nail holes, toe protection in the form of calks, and unsunk southern nail heads. Under a British influence that came into play with the Seven Years' War (1700 CE), a shoe was made from a bar with a groove and sunken nail heads, but no calks (Figure 8.1). Then, in England and the colonies in the 1800s, we see the coincidence of the invention of the mass-production horseshoe machine in the U.S. and the registration of farriers in the U.K. (Fleming, 1869). These two factors cement the lock-in, as we now have all the features of our modern shoe and all the elements/criteria for path dependence.

8.3 Path breaking and creation

We now return to the question of whether path breaking and path creation can be predicted or are even possible in the equine industry. Of the four sources of new variation and path-breaking in technology (Bassanini and Dosi, 2001), all are present in the horseshoe industry.

8.3.1 New actors, knowledge base, community of practitioners and forms of organization

We have described the shift in owners. As a result of this new shift, owners are now less knowledgeable about horses and less able to spend the time needed to provide complete care and keeping. This has created

a demand for support providers to help owners deal with their expensive and time-consuming hobby and to develop some horse savvy. In keeping with the sentiments of the new owners, much of this new industry adopts tools and techniques called natural horsemanship (NH), a term that did not exist before 1985 (McBane, 1992).[3]

A trainer named Pat Parelli wrote a book on natural horsemanship in the 1980s, which became a bible for a small group of western horsemen (Parelli, 1982; Parelli and Swan, 2005). As the new owners bought or boarded horses from such horsemen, the approach meshed with the sentiments of the new owner demographic. The eventual result can be considered a social movement (Rose, 1997; Rucht and Neidhardt, 2002). Scores of clinicians – basically celebrity horsemen with backgrounds in training and instruction each with their own brand of natural horsemanship – toured the country conducting clinics, selling videos/books/tools, and promoting social clubs dedicated to natural horsemanship. Parelli's Savvy Club is one example. Some sell out arenas, others work out of fairgrounds. All offer a philosophy to anchor a natural identity (for a leading example, see www.parelli.com). Of special interest is the barefoot segment, with its own set of leading practitioners, websites, books, and products.

Table 8.5 Where owners get information about horsemanship

Media source of reference (horseman)	**Per cent respondents**
websites	60
magazines	81
associations	50
clinicians	30

Table 8.6 What kind of information owners seek

Media sources	
Category	**Per cent sources**
Natural	25
Traditional	15
Association	30
Barefoot	20
Popular/ Mixed	30

Note: N=63 media sources named by 207 respondents.

A look at the media sources cited by owners provides a glimpse into the new knowledge base. In our survey, 60 per cent of respondents refer to websites and 30 per cent to clinicians for information on horsemanship (see Table 8.5). Magazines and associations are also frequently cited, but they may represent traditional or alternative knowledge bases. Taking a look at the 63 media sources identified by respondents, less than half would have existed 20 years ago. In addition, 50 per cent of our sample said they had one or more favorite clinicians. Of the 52 clinicians named, the vast majority practice natural horsemanship (see Table 8.6).

Clinton Anderson is one of the more popular natural horsemanship clinicians, who has the slogan 'No worries training for you and your horse'. That tag-line is important, as it again highlights the shift in ownership to folks who need training themselves, as well as for their horse. The homepage for Anderson's 'Down Under Horsemanship' features as a prominent image an ad for pink t-shirts emblazoned with 'precious'. This industry is not there to support the working rancher/horseman of the Old West, to be sure.

8.3.2 Heterogeneity among actors with imperfect adaptation

While we have described the new owners broadly in homogeneous terms, the reality is much different. The new owners share, in particular, the characteristic of what they are not: working ranchers. They do not share a common definition of who they are, however – they are young/old, simple/educated, blue-collar/white-collar, high-status/low-status, wealthy/middle-class, laborers/intellectuals, and so forth. A few, in fact, are skilled and experienced in showing and competing, while others come to horses almost entirely ignorant of equine matters. The unsophisticated new horse-owning demographic finds itself dependent for information, and hence identity, on the multitude of providers noted. Sources of social identity include anyone who can tell our owner how to be a horse owner, what a horse owner looks like, how they act, how they care for horses' hoofs. They embrace anyone who can help resolve their uncertainty, or fulfill their desire to be either natural or traditional. The identity of the horse owner takes on elements of those providers that help resolve the owner's uncertainty and dependence. Precisely because the horse owner seeks to be independent and individualistic, they seek to inherit aspects of providers' identities that would be needed to fulfill their own desired identity. That is, their identity would strive to take care of their horses by themselves, so they search for attributes needed to be self-made, stand-alone horsemen and women (Hoare, 1991; Cinnirella, 1998; Howard, 2000).[4]

Adaptation is imperfect in part because of this difference, but also in larger part because of the way information disseminates – through social networks. The typical new owner enters the horse world through contact with one of these providers, and then adopts that provider's referral network – a referral network that now directs beyond the traditional providers to other friends, clinicians, programs, and so forth.

Our survey reveals the heterogeneity in social network referents among our sample of horse owners. A majority rely in some part on their own thinking, as facilitated by the new knowledge base. Nearly half cite family and friends as influences, and a similar portion cite a trainer (see Table 8.7). Respondents often cite multiple sources, in varying combinations, which enhances idiosyncratic adaptation. Many of these new information sources question the functionality of the standard all-purpose steel shoe and prescribe alternatives ranging from modified shoes to no shoes at all, while others promote tradition.

The twin pulls on social identity created by the dependence of owners on others and the variety of influences on owner's decisions about hoof care is illustrated by our survey results (see Table 8.8). Nearly all in our sample see their farrier as influential, but are subject to the influence

Table 8.7 Who influences owners about horsemanship

Source of reference (horseman)	**Per cent respondents**
family/friends	46
own thinking	57
trainer	49
farrier	10
vet	12

Table 8.8 Who influences owners on hoof care

Source of influence (hoof care)	**Per cent respondents**
farrier	90
trainer	21
manager	5
equine vet	59
family/friend	10
associates	5
clinician	5
own thinking	36

of at least one other source. Veterinarians, trainers, managers, family, friends, associations, clinicians all received mention.

Let's take a quick look at a couple of these. Trainers and associations each pull on the social identity of the contemporary horse owner. The exact nature of the influence of associations depends on the organization, but there are some common points. The associations stress the social aspect of horse ownership: fulfillment, via competition, or through camaraderie; status, in many flavors, as some emphasize elitism, while others appeal to the common man; and tradition, according to a discipline, be it Western, English, or otherwise. The traditional trainer has an image that epitomizes the characteristics of the horseman: independent, individual, righteous. In general, private trainers tend to maintain control over all or most aspects of the care and keeping of horses under their training. Many traditional trainers portray new technologies and practices as unsafe for both the owner and the horse.[5] Through this kind of social influence, they seek to impart their own identity onto new owners.[6]

The equine vet is a professional with an emphasis on the medical knowledge base. Today, equine vets wear scrubs, with a clear and intentional message. In terms of the social identities of the vet and the horse owners, the vet 'does no harm' and thus attempts to reinforce the autonomy of the owner while providing appropriate care. To do so, the vet must read the owner for clues as to the source of their identity, and then frame the situation and alternatives accordingly. Doing so may seem to reinforce whatever identity the owner already has, but this is not entirely so. Since owners' identities are typically incomplete where their horses are concerned, the vet's role in resolving their dependence draws them to adopt aspects of the vet's identity as well.

8.3.3 Co-evolution technology and practice

As the developments in shoeing run parallel to those in horse training and usage, the technologies of shoes and hoof treatments evolve at the same time as practices for trimming and farriery and at the same time as knowledge and practice of horse training techniques, such as those promoted by the natural horsemanship and barefoot movements.

The ability of the farrier was seen as a key factor in owners' hoof care decisions and the farrier was taken-for-granted as an influence by respondents in our survey. The identity of the farrier is by and large, an anchor in tradition. There are 50 known farriery schools in the U.S., and just 15 certifiers that enforce a singular curriculum. Pictures on websites for the schools are interchangeable: the image of the strong,

hard-working, smithy is universal: hammer-swinging, anvil-pounding, wielding iron-tongs over a hot furnace. Farriers have been known to brag about the size of the foot they are able to achieve, which is done using the standard shoe. To the extent these farriers believe that a large foot provides better support for the horse, they will hold doggedly to tradition and will pass on such beliefs to owners.

Yet, farriers need to understand where to load the foot in order to trim the hoof and place the shoe. To do so, they must see the inside of the foot when looking at outside. Farriers that can do so are bothered by a long toe and run-under heel; a progressive farrier, then, needs to understand the biomechanics of the horse's hoof, but this sort of 'book-smarts' is not symbolically emphasized in the farriers' identity. Instead, it is revealed only if one is willing to dig past the front pages and into the fine-print of the curriculum. The switch from a hammer-wielding, anvil-pounding, iron-worker, to a professional care-provider concerned for the horse occurs through social influence and experiences. This occurs in three ways: i. by repeatedly being called out to correct a lameness issue caused by improper shoeing; ii. by apprenticing with an experienced, progressive farrier; and iii. by profound client demand. With regard to the latter, even a stubborn, traditional farrier may change approach if they are turning down a lot of work in order to maintain their identity.

Barefoot farriers, and others specializing in alternative methods, are entering the profession and partitioning the market – or realizing the value in the extant resource partitions provided by the changes in horse ownership and use. There is a clear segmentation between the barefoot and traditional farriers and their clientele. They are not competing directly. But, they are doing so indirectly – for influence over aspects of the social identity of owners and of farriery.[7]

A lot of new variation has emerged in the horseshoe industry as a result (Van Wyk, 2000). Several new varieties of boots/shoes/sandals are intended to replace steel shoes and be more comfortable for the horse. Plastic and rubber materials for shoes are another attempt to tap into the sentiments of the new owner demographic. They rely on the fact that these owners view their horses as part of the family, rather than livestock. There are two parts of this appeal to identity: humaneness and humanity. Humaneness asks how you can pound nails into your family member's feet. Humanity asks how you would like wearing steel-soled shoes. Surely rubber or plastic would be more comfortable, absorbing shock and providing some flexibility. Notice how these questions beg the owner to use their own thinking (asking how you…). In

addition, many are either recyclable or made from recycled materials, fitting with the natural identity of our owners. The rubber boot goes a step further, as no toxic glue is needed, and the humanity is obvious.

8.3.4 Invasion of new forms developed elsewhere

Here, we include new materials and designs, but also new ideas from the larger natural lifestyle movement. In the larger industry, we've earlier noted the new forms of clinicians touring the country, supported by horsemanship clubs, websites, and so forth. In addition, many of these clinicians are involved in cross-promotion with clothing lines and tack – you can even get your Parelli NH credit card (see www.parelli.com). For a broad array of products, endorsement by (or alignment with) a well-known clinic personality has become an important market differentiator. There is also a cable TV channel dedicated 24/7 to equine programming. These are all examples of forms that are new to equine industry, but are otherwise familiar to the new owners.

Focusing on materials and designs for shoeing, this part of the industry has exploded with the shift. The last decade saw hundreds of new firms, 600 different types of shoes, 50 kinds of nails and 30,000 new farriers (*Fortune*, 1993, 'Start ups add new look to $2 billion per year horseshoe business'). There is an annual sales base of $80 million for horseshoes within a $2 billion per year hoof care industry. New designs continue to date, but include synthetic materials, new metals, alternate shapes and configurations, and alternate technologies, such as hoof hardeners.

These variations all have either or both of the following invasive attributes. First, many were brought in by firms from other industries seeking to seize on opportunities. Rubbers/polymers, boots, materials, and hardeners are all borrowed. We cite an example in which 3M produced a shoe intended to replace the metal surface with an elastomer in order to reduce road damage (reported in *Design News*, 1989). It was later marketed to appeal to the new horse owner.

Table 8.9 How owners shoe horses

How shod	Percentages of horses
Round-toed steel	41
Round-toed alum/other	5
Square-toed	14
Glue-on	1
Boots	6
Barefoot	41

Second, many come in as stowaways on the natural bandwagon – appealing to the idea of a more comfortable, natural means of protecting the hoof. Synthetic materials, boots, and the Natural Balance shoe (square-toe shape) all fit in this category. The name Natural Balance by intent taps into symbolism of the larger movement. This is a patented design that allows the toe to wear off and the center part of the hoof sole (known as a frog) to contact the ground. These are two key features of the biomechanics of hoof movement in nature, which are not replicated with standard round-toe shoes.

The barefoot segment emphasizes the use of hardeners and boots. Owners influence each other, and focus on friends and family as referents. Family and friends tend to reinforce the lifestyle movement identity associated with the new owner demographic. They get their cues mainly from media sources that offer a natural, cutting-edge image of the horse owner.

In our survey, we see the impact of the new variation (see Table 8.9). While 41 per cent of the horses owned by respondents are reported to be shod using the standard round-toed shoe, an equal percentage are kept without shoes. The square-toed, Natural Balance shape, alternate metals, and boots each also have some toe-hold, so to speak.

8.4 Conclusions and directions

The evidence supports an intuition that identity matters for path persistence versus path creation, as the adoption of new variations maps onto the various agents of identity as social referents of horse owners in our survey. Table 8.10 recaps the key correlations among identity, influence, and adoption that have been woven into our exposition of path dependence and path creation. The equine industry suggests that unraveling the hidden dynamics of path dependence and creation requires

Table 8.10 What identities create which paths

How shod	Source of influence	Factors in how shod	Reasons for owning	Vocation
RT-steel	trainer/mgr	farrier's ability	hobby	
RT-alum	associates	performance	compete	horse related
NB	vet/clinician	what others do	not hobby	white-collar
boots	own thinking	own beliefs	companionship	ret., head
barefoot	family/friends	not performance	lifestyle	ret, hands
glue-on	some trainers		recreation	blue-collar

us to further incorporate the literatures on social identity, movements, and networks.

Notes

* Direct all correspondence to the first author at 1130 E. Helen Street, Tucson, AZ 85721-0108, USA kkoput@u.arizona.edu. We thank Elizabeth Hernandez for her invaluable assistance with all aspects of the project.

1. The survey was sent to a random sample of 500 horse owners in southern Arizona. Names were taken from client lists of local providers (vets, farriers) with an additional sample in Southern California. The survey asked demographics, reasons for horse ownership, media and personal sources of information/influence and factors considered on horsemanship and hoof care, group affiliations, names of providers, details on horses, how they are kept, used and shod. The sample yielded responses from 207 owners of 606 horses.
2. Personal experience of the second author.
3. The ideas of natural horsemanship have been traced as far back as the Greek writer Xenophon's treatise 'On Horsemanship', penned sometime between 400–350 PCE.
4. Take the following anecdote as an illustration of the interplay of identity and social influence. One of the authors, an equine practitioner, was struggling to get a horse owner to understand the cause of her horse's repeated problems. The practitioner suspected the horse's feeding was to blame. The diet included an array of supplements that were marketed variously as natural, holistic, and alternative. The owner was a rather bohemian woman – describable perhaps as a burned-out old hippie. Hence, the notion behind the feeds was appealing to her sense of who she was and how she constructed her own diet. When the practitioner reframed the attributes of the feeds as 'highly-processed junk,' the owner responded by switching to plain grass hay – resolving the horse's chronic problems.
5. One of the authors relates the reaction of his trainer to a new saddle he had purchased and brought to the stable. The saddle was alternative in design and did not conform to the Western saddle that she had recommended. The trainer repeatedly accused the owner of being unsafe and stated her intent to ban his saddle from the property.
6. Trainers go so far as to prescribe not only the type of tack an owner should use, but how an owner should dress – down to the brand of boots and jeans (in the Western case, britches for English).
7. As an example, the barefoot segment tries to claim trimming as a special domain. One respondent to our survey wrote in the margins on the form: 'I use a certified barefoot trimmer!' In reality, traditional farriers always trim as part of shoeing and any good farrier could be considered certified for barefoot trimming.

References

American Horse Council (2005) *National Economic Impact of the U.S. Horse Industry*. Washington, DC.

Arthur, B. (1989) 'Competing technologies, increasing returns, and lock-in by historical events', *Economics Journal*, vol. 99, pp. 116–131.

Bassanini, A. and G. Dosi (2001) 'When and how chance and human will can twist the arms of Clio', in R. Garud and P. Karnoe (eds) *Path Dependence and Creation*, London: Lawrence Erlbaum, pp. 41–68.

Bowker, R.M., K.K. Van Wulfen, S. Springer and K.E. Linder (1998) 'Functional anatomy of the cartilage of the distal phalanx and digital cushion in the equine foot and a hemodynamic flow hypothesis of energy dissipation', *American Journal of Veterinary Research*, vol. 59, pp. 961–968.

Bowker, R., A.M. Brewer, K.B. Vex, L.A. Guida, K.E. Linder, I.M. Sonea and A.W. Stinson (1993) 'Sensory receptors in the equine foot', *American Journal of Veterinary Research*, vol. 54, pp. 1840–1844.

Cinnirella, M. (1998) 'Exploring temporal aspects of social identity: The concept of possible social identities', *European Journal of Social Psychology*, vol. 28, pp. 227–248.

Clark, B. (1831) *An Essay on the Knowledge of the Ancients Respecting the Art of Shoeing the Horse, and the Probable Commencement of This Art*. London: The author.

David, P. (1985) 'Clio and the economics of QWERTY', *American Economics Revue,* vol. 75, pp. 332–372.

Design News (1989) 'Engineered horseshoe protects country roads', vol. 45, p. 112.

Ensminger, M.E. (1990) *Horses and Horsemanship: Animal Agriculture Series*, 6th edn. Danville, IL: Interstate Publishers.

Ensminger, M.E. (1991) *Horses & Tack: A Complete One Volume Reference on Horses and Their Care,* Rev. edn. Boston: Houghton Mifflin.

Evans, J. (1990) *The Horse*, 2nd edn. New York: Freeman.

Fefer, M. (1993) 'Horses need shoes too', *Fortune*, December 27, 1993, vol. 128 (16), p. 'Job Tip'.

Fleming, G. (1869) *Horseshoes and Horseshoeing: Their Origin, History, Uses and Abuses*. London: Chapman & Hall.

Gartland, M.P. (2005) 'Interdisciplinary views of suboptimal outcomes: Path dependence in the social and management sciences', *The Journal of Socio-Economics,* vol. 34, pp. 686–702.

Heymering, H. (2008) 'Who invented horseshoeing?' http://www.horsehoes.com, date accessed 4 April 2007.

Hoare, C.H. (1991) 'Psychosocial identity development and cultural others', *Journal of Counselling and Development,* vol. 70, pp. 45–53.

Howard, J.A. (2000) 'Social psychology of identities', *Annual Review of Sociology,* vol. 26, pp. 367–393.

Liebowitz, S.J. and S.E. Margolis (1990) 'The fable of the keys', *Journal of Law and Economics,* vol. 22, pp. 1–26.

Liebowitz, S.J. and S.E. Margolis (1995) 'Path-dependence, lock-in and history', *Journal of Law, Economics, and Organization,* vol. 5, pp. 205–226.

Mahoney, J. (2000) 'Path dependence in historical sociology', *Theory and Society,* vol. 29, pp. 507–548.

McBane, S. (1992) *A Natural Approach to Horse Management*. London: Methuen.

Miel, R. (2005) 'Well heeled: Rolco takes PU product into big horseshoe market', *Rubber & Plastics News,* July 25, 2005, p. 'News'.

Parelli, P. (1982) *Natural Horse-Man-Ship*. Western Horseman Books.
Parelli, P. and K. Swan (2005) *Raise Your Hand if You Love Horses: Pat Parelli's Journey from Zero to Hero*. Western Horseman Books.
Rose, F. (1997) 'Toward a class-cultural theory of social movements: Reinterpreting new social movements', *Sociological Forum,* vol. 12, pp. 461–494.
Rucht, D. and F. Neidhardt (2002) 'Towards a movement society? On the possibilities of institutionalizing social movements', *Social Movement Studies*, vol. 1, pp. 7–30.
Severin, T. (1989) 'Retracing the first crusade', *National Geographic,* vol. 176, pp. 326–365.
Van Wyk, A. (2000) 'Making the shoe fit/modern touches to an old trade', *The Calgary Sun*.
Zippelius (1891) 'Historical development of the horseshoe', *Scientific American*, vol. 819, pp. 32–819.

9

Path Dependence in Merchant Institutions: The Case of French Public Timber Sales

Gérard Marty

9.1 Introduction

The rules for selling timber originating from French public forests are set by a specific law, which in the course of history has granted a central role to the auction system. For centuries, and until recently, this sales mechanism was the only one that ensured the distribution of timber. However, due to the lack of competitiveness in the French timber industry the auction – as an institution – (North, 1994, p. 360) has been blamed in the last few decades in several reports as being responsible for the difficulties suffered by timber professionals. Even so, the auction still remains the favorite mechanism for selling public timber. This situation leads us to question the reasons for the persistence of this institution over a long period of time.

To help us answer this question 50 interviews were conducted in 2007 with buyers (loggers, sawyers, and manufacturers of paper and timber panels) and the staff in charge of selling timber at the French Public Forest Service (Office National des Forêts, hereafter ONF).

The next section describes the evolution of the law system in order to better understand France's adherence to the auction system for selling French public timber (essentially standing timber lots).

Section 9.3 reviews the main characteristics of the concept of path dependence. The concept is then applied to the case of French timber sales. I argue that a self-reinforcing process due to the presence of learning effects explains the reproduction of the auction system to sell public timber.

In addition, I explain how the institutional lock-in of the timber sales mechanisms has affected the organizational shape of the French timber industry.

9.2 The historical predominance of the auction

I cannot fully understand the public timber sale mechanisms without a preliminary study of their historical evolution. In this regard, Colbert's ordinance[1] is the reference text for timber sales (Estrade and Morin, 2006). The ordinance made the auction the merchant institution which was able to sell public timber efficiently.

9.2.1 From the ordinance of 1669 to the first Forestry Act: the omnipresence of the auction

Title XV of the ordinance describes the selling modes of the King's timber. The auction may be regarded as a formal rule, which aimed at restoring some transparency concerning timber sales in order to fight local collusion. The ordinance was so influential that two centuries later when the legislators composed the first Forestry Act (1827) they maintained the auction to ensure that no fraud took place during timber sales.

In fact, auction sales provide an additional guarantee against the risk of wrongly estimating the value of a timber lot. Indeed, the numerous competing buyers gathering at the same place at the same time tend to approach the 'true market value.' The auction system gives more security to the seller against false estimation than posting a fixed price in advance.

9.2.2 In the twentieth century: the erosion of the monopoly accorded to the auction

The monopoly held by auction sales for centuries was first weakened in 1926. For the first time in centuries, some public forest timber was allowed to be sold under tight conditions by private agreements. However, the auction remained the main selling practice. Sales by private agreement were limited to certain types of wood that were not sold at the auction (burned and damaged wood).

An historical analysis of legislative changes throughout the twentieth century shows that certain institutional inflexibility exists concerning the selling practices of timber. Although the Forestry Act was slightly transformed in 1969 to allow sales by private agreement as a derogation, the auction remained the rule. This explains why in 1982 the Duroure report stressed the necessity to propose new selling practices to secure the future of the timber industry. In 1998, the Bianco report reiterated that there was a problem related to the competitiveness of timber prices in France due to the organization of timber sales. Bianco showed that as a result of the random character of the auction system buyers needed to

estimate five to ten times more standing timber lots in the forests than they wanted to buy. These prospecting costs wasted a great deal of time and energy. Bianco proposed to modernize selling practices, but the auction continued to be the preferred system.

9.2.3 The beginning of the twenty-first century: the end of the monopoly held by the auction

Whereas the different amendments to the Forestry Act during the twentieth century gave a limited position to the private agreements while preferring the auction system, the beginning of the twenty-first century places private agreements and the auction system in a tie position from a legal point of view.

The situation described in the reports was taken more seriously from 2001 onwards. That year, for the first time in centuries, the law on forest orientation allowed sales practices other than the auction to be used in situations not covered by a derogation. However, since the application decrees were not properly composed, the use of sales by private agreement remained rare in practice and the difficulties related to supply persisted.

In 2003 the Juillot report insisted that these supply problems were caused by the sales of small-sized heterogeneous lots at auction. Indeed, this mechanism leads to high direct costs (logging and transport) and a costly security stock due to the lack of visibility in timber supply (Juillot, 2003, p. 37). It also recommended more contractualization in sales by private agreement (Juillot, 2003, p. 40). The law of 23 February 2005 finally implemented in Article 134–7. From now on, the people at the ONF in charge of selling timber are free to choose between auction sales and sales by private agreement. Eighty years after the appearance of the first derogation, sales by private agreement became an independent sales practice.

Nevertheless, as I will explain in the next section, the massive implementation of supply contracts through private agreements planned in the State-ONF for 2007–11 contract is already provoking opposition from professionals in the timber industry.

However, it is too early to say how the timber sales mechanisms will evolve in the future. Indeed, the ONF has had the possibility to choose between auction and private agreement for only one year.

9.3 Path dependence and the institutional lock-in into the auction

The historical use of the auction for several centuries has developed path dependence in the specific rules created for selling public timber.

The path represented by the auction has resulted in the stability of the merchant institution, which has led to a situation of inefficiency for a section of the timber industry.

9.3.1 The auction: path dependence and the timber industry structuring

The concept of path dependence was first used to explain the influence of minor historical events (Arthur, 1989) on the future, concerning the development of some technological choices. In the timber industry, this concept allows one to understand why the auction has remained the fundamental institution for the sale of timber and how it has played an active part in its shaping.

9.3.1.1 The origin of the concept of path dependence

Taking into consideration past decisions in order to understand present choices and future projects is related to an historical perspective on social sciences. But for orthodox economic reasoning, the fact that economy is not 'embedded' in the society, made possible by a regulated merchant area – has erased the impact of history in the economic analysis (Abolafia, 1996).

Contesting this un-historical view, new approaches arose trying to give importance to history to explain economic situations and their evolution. These studies, which are in line with the concept of path dependence, underline that history matters (David, 2007, p. 92). They pursue a perspective of retracing the history of events in order to understand today's choices (Berman, 1998). The first studies referring to this concept talked about technology choices. Nevertheless, it is also possible to apply it in the case of timber sales. To this end, let us present what are the constituent elements of a path dependent process.

Although there is no distinctive definition, I can consider that the concept of path dependence refers to a dynamic process where final outcomes depend on historical events, which are not purely random. Once a particular option has been selected from several possibilities, it becomes difficult to reverse this choice as the original causes have disappeared. This fact could be explained by the presence of sunk costs (Balmann et al., 1996) or self-reinforcing processes (Mahoney, 2001), also called 'increasing returns' by economists (Pierson and Skocpol, 2002).

These points contrast with the neoclassical analysis, which neglects the role of history and considers that the process of economic allocation leads to a unique efficient equilibrium (Puffert, 1999). As Paul David

(1985, p. 332) has shown, a path-dependent sequence of economic variations could be the result of chance elements rather than systematic forces and does not necessarily involve an efficient solution. Brian Arthur (1989) has also underlined this possibility and pointed out that because of small events the economy could be locked in into an inferior technology. Now that I have defined the concept of path dependence, let us suggest a new field of application for it: timber sales.

9.3.1.2 The transposition of the concept of path dependence into timber sales

As previously noted, the concept of path dependence has often been used to explain the dominance of inferior technologies (David, 1985; Arthur, 1989; Cowan, 1990; Puffert, 2002). However, an increasing number of studies referring to this concept have also appeared in other areas, such as organizations (Hunt and Morgan, 1996; Bruggeman, 2002; Staber and Sydow, 2002) or institutions (Roe, 1996; Pierson, 2000; Mahoney, 2001; Hathaway, 2001). In the case of institutions, the studies mentioned above have emphasized how the concept of path dependence is useful for interpreting their specific trajectories.

In practice the study of institutions describes an historical sequence wherein the game rules generate a self-reinforcing process (North, 1990). This means that I first need to describe the historical events which have caused the selection of one option among a range of possibilities. The next step then consists of emphasizing – once a choice has been made between alternative solutions (critical juncture) – that, due to self-reinforcing process, it is difficult for actors to reverse course (Pierson, 2000). The presence of a self-reinforcing process leads to certain rigidity, which can lock in future choices. This approach requires us to take into consideration the power of actors' position. Indeed, it should provide explanations about the conflict of interests, which can appear between actors because of unevenly distributed costs and benefits (Mahoney, 2001, p. 114).

In the case of the sales of French public timber, our historical approach leads to consider that specific events in the past centuries still have an impact on the evolution of the sales mechanisms and the structure of the timber industry. Originally, the auction was institutionalized in order to limit fraud during sales. Nevertheless, three centuries later, fraud having been almost eliminated, it is still the main sales mechanism. The persistence of the auction, despite heavy criticism, leads us to question whether the legislation in the field of timber sales is path dependent. This path dependence could be explained by the increasing

returns obtained by members of the timber industry who in time learned the specificities of the auction mechanism.

In practice, the stability given by keeping the same sales mechanism for several generations has facilitated the appearance of a learning-by-using effect (Arthur, 1989, p. 116) in the oral descending auction. The specific character of the descending auction mechanism – where an auctioneer announces different prices at a fast rate – requires experienced staff because of the high rhythm of the sales. According to the buyers, learning how to be in control of the descending auction can take many years. Moreover, even with the necessary level of experience this sales mechanism is referred to as being stressful and dangerous.

Besides, using the auction over a long period of time improves the buyers' knowledge about the needs of the other professionals. Indeed, comparing the sales results of the current year with the ones of previous years is fairly easy since the sales calendar is about the same every year and the number of participants varies very slightly.

The fact that the buyers know each other allows them to guess who will make a bid, depending on the lot, and to anticipate the level of competition. This situation is particularly true in the descending auction because it is necessary to be physically present in the room to buy lots. However, in the sealed-bid first-price auction, which has been more widely used since 1989, it is feasible to make a bid by mail and thus create doubt on the level of the competition.

Besides, buyers are mostly family-owned companies where the auction mechanisms are explained from generation to generation to obtain optimal results (Barnes et al., 2004). Buyers have mentioned that they learned the auction mechanism through their recurrent presence at sales, starting when they were young.

Finally, the skills developed by the buyers concerning their competitors' bids lead them to share out the lots tacitly when the quantity proposed is weighty. As each buyer knows exactly the supply locations of all their competitors, the lots are easily divided. The auction is not only a sales mechanism but represents a way to act, which has been integrated into the participants' habits through an intergenerational learning process.

The existence of those skills – gained as a result of several years of practice or even intergenerational experience – explains why some professionals wish to keep the auction and may feel threatened by the new possibility of selling timber by private agreements.

As regards the seller, increasing returns are possible due to the skills compiled during many sales organized by agency chiefs at the ONF.

Order in the catalogue, rhythm of sales, management of the reserve price – all these are elements, which allow the sales director to influence the results of the auctions. Moreover, like buyers, agency chiefs can estimate the level of competition for some lots, based on the results recorded in the previous years. All this information leads to a domestication of this sales mechanism and contributes to the development of a social inertness around the auction. In this situation only a major event could make the agents change the sales protocol. In the case of timber sales, this event took place during the storm in 1999. The large quantity of fallen timber put on the market on this occasion obliged the ONF to use private agreements in order to sell the timber faster. This fortuitous event has allowed the development to a larger scale of a sales mechanism normally rarely used. Furthermore, in the Lorraine region, this situation coincided with the arrival of new sales staff. The latter had no auction experience and did not benefit from the learning effect like some of their colleagues.

This distinction now has a considerable impact on the attitude of the ONF staff towards the growing use of supply contracts through private agreements, which was the purpose of the law of 23 February 2005. On the one hand, there are agents with experience of the auction mechanism who are still not very confident about the conditions of supply contracts and the risks linked to their limited knowledge about private agreements. On the other hand, I find agents who seem less dependent on the auction system and more open to applying the institutional change.

The influence of past events is clearly present in the choices concerning the sales mechanism. Nevertheless, after considering that agents could historically spark off an institutional change, it is also important to take into consideration the influence the auction has on the organization of the timber industry.

9.3.2 Institution and timber industry organization

The study of institutions is not limited to strictly identifying the original reasons of their appearance in order to analyze their consequences. The linear link 'cause and effect' is not enough to understand the reality of an institution and its context because effects also play a hand in modifying the reasons in a cumulative process (Veblen, 1909). Consequently, I should consider the impact of the setting up and the retaining of the auction, which has lasted more than three centuries, on the organization of the timber industry. The provision of heterogeneous timber lots and the development of an intermediary network in the upstream of

the timber industry are the consequences of the fact that the auction has persisted as the mechanism for selling timber.

Unlike several countries (Germany, Austria, England), which sell harvested timber, in France a major role has been given to the sales of standing timber. Selling standing timber has a drawback because of the presence of heterogeneous lots (mix of hardwood and softwood species). This sales mechanism does not allow timber to be selected by categories of timber before the sales. Consequently, standing timber sales limit the possibility to use private agreements for a specific quantity and species of timber. Indeed, the supply contracts through private agreements – which concern a species and a quality of timber – are about harvested timber (except a few homogeneous lots meant for pulpwood).

Because of the heterogeneity of standing timber lots the forest legislation has remained focused on the auction. However, this 'cause and effect' link between the timber supply (standing timber) and the sales mechanism (auction) must not be interpreted in a strictly linear manner. In reality – as mentioned above – the context has structured the institution which guarantees the timber sales. The institution has also conditioned the environment in which it takes place.

Therefore, the auction – first considered as a consequence linked to the standing timber sales – has determined the shape taken by the timber industry during the supply. The almost exclusive use of the auction has finally led the ONF to concentrate its action on the management of forests and left the timber allotment to the backing of the timber industry.

To better understand the impact of the auction on the allotment of timber and its consequences on the organization of the timber industry, it is interesting to compare the role played by the ONF until now, as well as its future missions linked to the development of supply contracts through private agreements.

In terms of supply contracts, the buyers and the ONF determine the quality of timber sold by private agreements in advance. The buyers obtain a homogeneous lot of timber constituted by one species and one quality of timber. Their choices are made depending on the opportunities and the specificities of their production machinery. The buyers can refocus on their specific main field, that is the timber processing. Thus, the selection work of timber in forests is devolved to the ONF according to the criteria listed in the contract. The engagement of the ONF in the buyer's supply needs is even bigger.

Until now, taking into consideration the buyer's needs was less apparent during the constitution of the timber lots. The composition of homogeneous lots was not a priority for the ONF which sells timber through

the auction system. The ONF was simply not interested in promoting the allotment work before sales. Indeed, in the auction system, the role of the ONF was for a long time limited to choosing the timber to be harvested, with regards to the principles of a reasonable silvicultural management. In this configuration of heterogeneous standing timber sales, buyers have to select timber in the lots according to their production machinery and find new buyers for the rest of the lot.

This position in the timber industry has favored loggers and participated in the vertical integration between loggers and sawyers. The massive use of standing timber sales meant that processing firms had to find loggers in order to cut, haul, and carry timber to the factory. But beyond this work, loggers appear incontrovertible to the timber industry. Due to their knowledge of the timber industry's needs and know-how to enhance the value of timber lots, loggers became the agents who are in charge of the timber allotment.

Many professionals consider loggers to be the main characters in supplying the timber industry. Loggers have developed skills to enhance the value of standing timber lots, which leads them to give value to the whole lot. This is more often in the case of hardwood where the variety of uses of a log leads to the allotment of portions according to the activities: peeled veneer, slicing, sawing. Due to the constitution of buyer networks beyond French borders, loggers have guaranteed that all the timber is sold in spite of heterogeneous lots.

The presence of loggers as the link between forest production and its processing has been reinforced with the auction. Indeed, in the auction sales the risk of not obtaining lots obliges the buyers to visit numerous lots in the hope of acquiring enough lots for their activity. As a consequence of this high cost of estimating lots and the risk of not obtaining the visited lots during the auction sales, many industries have transferred this job to the loggers.

The auction has finally led to the faint possibility of the ONF giving value to timber lots in favor of loggers. Indeed, the sales director has little information during the sales concerning the future use of the timber sold. Consequently, the administration is deprived of crucial information to enhance the value of timber.

The auction – the consequence of a tradition of selling heterogeneous standing timber – has consolidated the central role of loggers in the timber industry supply because of a cumulative causality process. In such a case the new law adopted in 2005, allowing the ONF to freely use private agreements seems like a factor that reconsiders the stability of the timber industry's structure. The will of the ONF to take back a

part of the enhancing action of timber with the rise of supply contracts through private agreements is causing a strong resistance to this change in the timber industry, particularly from the loggers making timber transactions since they are excluded from this sales mechanism.

9.3.3 Lock-in effect in timber sales

In the classical evolutionary paradigm it is considered that whatever survives is efficient (Friedman, 1953). Thus, laws or institutions which are not efficient quickly find themselves competing with more competitive solutions which should supplant them (Posner, 1993).

However, the analogy made with the evolution of species in biology is not obvious because people have a better capacity than plants to take part in changes throughout history (Roe, 1996, p. 665). Unlike plants characterized by a limited capacity to constrain their environment, people have the capacity to modify institutions, which are the result of a social process (Bellon, 2002).

Moreover, it is feasible – referring to some historical events – to find situations for which the economic actions taken have caused the adoption of a rule that will lock the market and limit the rise of innovation or new rules. This kind of situation is characterized by a growing rigidity in choices as the institution is adopted and used by the participants. But this rigidity risks obstructing the development of solutions that could be superior, while maintaining an inferior system. This fact puts into doubt the principle of the optimum equilibrium.

In public timber sales the existence of a sub-optimal situation for a part of the timber industry could be explained by an institutional lock-in, which is based on the auction system. A proof of this is the criticism by the representatives of the processing industry who have big timber needs and who criticize the auction for their lack of visibility concerning the quantity obtained during the sales (Juillot, 2003).

This point is not new, as it is found in several reports about timber: Duroure (1982), Bianco (1998), Juillot (2003). They insist on the difficulties linked to supply modes, which are the result of the auction being the primary mechanism to sell timber. The repetition of this analysis in different periods is significant of the social inertness created by the institution (Chavance, 2007, p. 20).

As it was well adapted to a local network constituted by family-owned sawmills, which require only small quantities of timber, the auction did not cause supply problems until the 1980s and seemed to be suitable to the timber industry (Badré, 1984). But due to the concentration of firms, the increase of the processing industries' working capacities

and the appearance of international competition, the auction system was contested. Indeed, big processing units have important sunk costs and demand high visibility on prices and quantities available over long periods of time. Unlike the supply contracts through private agreements the auction system does not guarantee any quantity or price to the bidders.

Thus, the formal rule adapted to the past became inefficient to regulate the present situation. The institutional lock-in of the timber sales results in a lack of competitiveness for the timber industry.

9.4 Concluding remarks

The sales mechanism used to sell timber is not the result of a simple economic 'cost – advantage' trade-off at a moment in time, but rather constitutes a social construction in which the weight of past decisions has a significant influence on future choices.

Thus, the auction – as an institution – has contributed to the organization of the timber industry's structuring over more than three centuries. This stability in the sales mechanisms has provoked the apparition of path dependence in the auction system. Maintaining the auction system has allowed increasing returns due to learning effects. The training methods of the auction system are passed down from generation to generation creating a behavioral lock-in. Each step made on the same path due to habit has reduced the cost of a new step on the same path, creating a self-reinforcing process around the auction. With time, this path dependence in the auction system has led to an institutional inertness, inducing a loss of competitiveness for a section of the timber industry.

However, the present evolution of the power struggle between economic and political special-interest groups (Fligstein, 2001) inside the timber industry seems to be changing the organization of sales mechanism. It seems possible to consider that instead of having a path breaking process (Windrum, 1999), the institution of the old path continues and a new institution emerges (Deeg, 2001, p. 7).

Notes

1. Colbert was the French Minister of Finance under King Louis XIV.

References

Abolafia, M.Y. (1996) *Making Markets: Opportunism and Restraint on Wall Street*. Cambridge: Harvard University Press.

Arthur, B.W. (1989) 'Competing technologies, increasing returns, and lock-in by historical events', *Economic Journal*, vol. 99, pp. 116–131.
Badré, M. (1984) *Forêt et Marché du Bois*. Paris: Hatier.
Balmann, A., M. Odening, H.P. Weikard and W. Brandes (1996) 'Path-dependence without increasing returns to scale and network externalities', *Journal of Economic Behavior & Organization*, vol. 29, pp. 159–172.
Barnes, W., M. Gartland and M. Stack (2004) 'Old habits die hard: Path dependency and behavioral lock-in', *Journal of Economic Issues*, vol. 38, pp. 371–377.
Bellon, B. (2002) 'Qu'avons-nous appris de l'évolution?' ADIS Working Paper, Université de Paris sud.
Bianco, J. L. (1998) 'La Forêt une chance pour la France: Rapport au Premier ministre', *La Documentation française*, Paris.
Bruggeman, D. (2002) 'NASA: A path dependent organization', *Technology in Society*, vol. 24, pp. 415–431.
Chavance, B. (2007) *L'économie Institutionnelle*. Paris: La Découverte.
Cowan, R. (1990) 'Nuclear power reactors: A study in technological lock-in', *Journal of Economic History*, vol. 50, pp. 541–567.
David, P.A. (1985) 'Clio and the economics of QWERTY', *American Economic Review*, vol. 75, pp. 332–337.
David, P.A. (2007) 'Path dependence: A foundational concept for historical social science', *Cliometrica*, vol. 1, pp. 91–114.
Deeg, R. (2001) 'Institutional change and the uses and limits of path dependency: The case of German finance', MPIfG Discussion Paper.
Duroure, R. (1982) 'Propositions pour une politique globale forêt-bois. Rapport Duroure', *Revue Forestière Française,* Special Issue.
Estrade, A. and G.A. Morin (2006) 'Historique de l'évolution du cadre législatif et réglementaire des modes de ventes des bois des forêts publiques', *Revue Forestière Française*, vol. 58, pp. 245–256.
Fligstein, N. (2001) 'Le mythe du marché', *Actes de la Recherche en Sciences Sociales*, vol. 139, pp. 3–12.
Friedman, M. (1953) *Essays in Positive Economics*. Chicago: The University of Chicago Press.
Hathaway, O.A. (2001) 'Path dependence in the law: The course and pattern of legal change in a common law system', *Iowa Law Review*, vol. 86, pp. 101–165 (Boston University School of Law Research Paper).
Huntand, S. and R. Morgan (1996) 'The resource-advantage theory of competition: Dynamics, path dependencies, and evolutionary dynamics', *Journal of Marketing*, vol. 60, pp. 107–114.
Juillot, D. (2003) 'La filière bois française: La compétitivité, enjeu du développement durable', *Revue forestière française*, vol. 3, pp. 185–296.
Mahoney, J. (2001) 'Path-dependent explanations of regime change: Central America in comparative perspective', *Studies in Comparative International Development*, vol. 36, pp. 111–141.
North, D.C. (1990) *Institutions, Institutional Change and Economic Performance*. New York: Cambridge University Press.
North, D.C. (1994) 'Economic performance through time', *American Economic R*eview, vol. 84, pp. 359–369.
Pierson, P. (2000) 'Increasing returns, path dependence, and the study of politics', *American Political Science Review*, vol. 92, pp. 251–267.

Pierson, P. and T. Skocpol (2002) 'Historical institutionalism in contemporary political science', in I. Katznelson and H.V. Milner (eds) *Political Science: State of the Discipline*. New York: W.W. Norton, pp. 693–721.

Posner, R.A. (1993) 'What do judges and justices maximize? (the same thing everybody else does)', *Supreme Court Economic Review*, vol. 3, pp. 1–41.

Puffert, D.J. (1999) 'Path dependence in economic history', in C. Herrmann-Pillath and M. Lehmann-Waffenschmidt (eds) *Handbuch zur evolutorischen Okonomik*. New York: Springer-Verlag Heidelberg, pp. 1–7.

Puffert, D.J. (2002) 'Path dependence in spatial networks: The standardization of railway track gauge explorations in economic history', *Explorations in Economic History*, vol. 39, pp. 282–314.

Roe, M. (1996) 'Chaos and evolution in law and economics', *Harvard Law Review*, vol. 109, pp. 641–668.

Staber, U. and J. Sydow (2002) 'Organizational adaptive capacity: A structuration perspective', *Journal of Management Inquiry*, vol. 11, pp. 408–424.

Veblen, T. (1909) 'The limitations of marginal utility', *Journal of Political Economy*, vol. 17, pp. 620–636.

Windrum, P. (1999) 'Unlocking a lock-in: Towards a model of technological succession', *Research Memoranda 010*, Maastricht: MERIT, Maastricht Economic Research Institute on Innovation and Technology.

10

Path Dependence, Lock-in, and the Emergence of Clusters: The Case of Istanbul's Film Cluster

Özlem Öz and Kaya Özkaracalar

10.1 Introduction

Film making, arguably the creative industry with the biggest cultural as well as economic impact, is considered both a craft and an industry, and the available evidence suggests that this form of economic activity tends to be concentrated geographically. Yet, there has been very little work examining the film industries in regions other than Hollywood (Dudrah, 2006). On the other hand, although the literature on clusters has been flourishing with interesting contributions in recent years (Martin and Sunley, 2003), it is rare to find studies in the existing literature that specifically focus on and systematically analyze the genetics of such agglomerations (an example of the exceptions is Sydow et al., 2010).

The present study aims to contribute to this literature by documenting the origins and the early historical geography of the motion picture industry in Istanbul with the ultimate purpose of advancing our current theoretical knowledge as regards the exact functioning of the process of path building in general, and the way it works in the emergence of clusters in particular. Specifically, this chapter digs into the ways of how lock-in logic operates, stressing the importance of the process rather than the outcome, for we can perhaps have a greater control over the outcome if we can grasp how it comes to be produced and reproduced.

This chapter is organized as follows. The following section reviews and discusses the major theoretical approaches in the literature that

have been put forward to explain the emergence of clusters, also relating them to the debates revolving around the concept of path dependence. The evidence available in this regard on world's major film-making centers is integrated into these discussions whenever possible, as the specificities of Istanbul's Yesilcam (the historical center of film making in Istanbul) will be better understood when examined with reference to the experiences of other film clusters. The chapter then proceeds to an in-depth analysis of the emergence of Istanbul's film industry. The theoretical implications revealed by the analysis conducted are discussed in the concluding section.

10.2 Emergence of clusters and path dependence: theory and evidence

If we briefly summarize the historical antecedents of cluster thinking, we should start with Alfred Marshall (1949, first published in 1890), who argued that specialized labour and inputs, local demand conditions, availability of related and supporting industries, information spillovers, and the ease of comparison and the resulting savings in terms of time and search costs on the customers' part are among the prime reasons why specific industries tend to be geographically concentrated. Marshall (1949) writes about the particular 'atmosphere' that characterizes such districts and believes that the mystery of such concentration is hidden in the very nature of this atmosphere. If, for instance, a clustering of economic activity like Hollywood is characterized by such an atmosphere as defined by Marshall (1949), it follows that if one wants to be in-the-know s/he might prefer 'to be there' since the advantages associated with it are difficult to tap into from a distance. Besides, it is very hard, if not impossible, to imitate the same 'system' in another locale, which in turn reinforces the likelihood of sustainability for the cluster (Öz, 2004). Following Marshall's pioneering ideas, other issues such as social capital, tacit knowledge and trust have begun to enter into the analysis in time and considered to play central roles in shaping the development patterns of different forms of 'sticky places' (Markusen, 1996), including 'learning regions' and 'innovative milieus' (Brusco, 1996; Storper, 1999).

In sharp contrast to the richness of theoretical approaches explaining the phenomenon of clustering, there is surprisingly little in the available literature by way of systematic empirical analysis of how and why a cluster emerges in a particular location (Scott, 2005). The existing empirical evidence suggests that the main factors triggering the

emergence of clusters might be diverse, ranging from the presence of 'special inputs' in a locale to pure 'historical accidents' (Krugman, 1991). Other scholars (e.g., Enright, 1990; Porter, 1990, 1998) have emphasized that the beginnings of a geographic cluster can often be traced to an unusually 'sophisticated local demand', the prior existence of 'related industries' and/or one or two 'innovative companies.'

It has been argued that the processes associated with the emergence and subsequent development of a cluster might have different economic logics. Accordingly, agglomeration economies and other external effects arise almost naturally after a cluster has taken off, whereas it is usually 'old economy' factors like firm-building capabilities, managerial skills, a substantial supply of skilled labor and connection to markets, which are crucial for the initial take off (Bresnahan et al., 2001). In a similar vein, it has been stressed that 'plain luck' is not enough in itself for the emergence of new clusters and that it needs to be given life by the spark of entrepreneurship, which in turn should be supported by social capital, venture capital, and entrepreneurial expertise and support systems (Feldman, 2001).

There is limited but steadily growing literature on the emergence of film clusters. One recurrent theme in this literature relates to the prominence of socio-spatial networks in film industries, reinforcing their tendency to concentrate geographically. According to Storper and Venables (2004, p. 351), face-to-face contact, which is an efficient communication technology facilitating information spillovers as well as providing psychological motivation is especially important in situations where information is rapidly changing and not easily codified which are in turn key features of many creative activities. Knowledge spillovers in particular have been argued to be highly dependent on face-to-face interactions and are thus expected to be highly localized and to decay rapidly with distance (Aharonson et al., 2007, p. 90). All this in turn favors long-term-oriented, highly personal relations in the film industry, and there might be serious barriers under these circumstances to developing relationships for those who are not members of these networks.

Several scholars have, on the other hand, underlined the impact of governments,[1] transnational corporations,[2] technological breakthroughs, and other external shocks[3] on the emergence, geography, and organizational structure of film clusters. It should also be stated, at this juncture, that many studies do, in fact, tend to emphasize the necessity of a simultaneous presence of various factors, which are theorized to be jointly determining the particular path taken by a cluster (Coe,

2000; Scott, 2000). This is in line with the co-evolutionary approach (Aldrich, 1999; Lewin and Volberda, 1999), whose application to the case of Hollywood (Jones, 2001, p. 912) has revealed that studying how institutional rules, firm capabilities, entrepreneurial careers and competitive dynamics have co-evolved in the early years of this cluster can improve our understanding of its emergence considerably.

The debate in the literature as to the specifics of how Hollywood, which is seen as one of the most remarkable examples of a successful industrial agglomeration in the world, has emerged provides additional insights. It has been argued that the principle reasons that drew moviemakers to the Los Angeles area include California's warmer climate (and light), its topography (diversity), and some economic concerns (e.g., Los Angeles was known to be a non-union city) (Lukinbeal, 2002, p. 251). Given that Southern California does not have a monopoly on sunshine and varied landscapes, as Scott (2005, p. 15) reminds us, the explanatory power of such accounts remains rather limited.

The path-dependence and lock-in perspective (Arthur, 1985; David, 1985) might be of help in this regard. In fact, both Arthur and David themselves have used geographical examples while developing the concept of path dependence (Martin and Sunley, 2006). Accordingly, many social processes may start off as purely random events but, if agglomeration economies are at work, the choice of location for additional units entering the system will become less and less random. There is a need here to be explicit about how a particular path emerges and how one technology may then evolve to dominate the market and become locked-in[4] as the industry standard, given that this process might be rather complex, involving, for instance, equifinality (many different paths to the same outcome) as well as multifinality (many different outcomes from the same value of an independent variable) (Bennet and Elman, 2006, p. 251).

It should be noted, however, that the path-dependence perspective itself is not immune from criticism. The approach is, for instance, criticized for often attributing the inefficiencies that emerge as a result of lock-ins to 'historical accidents.' Doing so indeed minimizes the role firms and entrepreneurs play in creating and shaping their environment. Seeing firms and entrepreneurs not as passive respondents but as influencers and shapers of their environments in turn makes it possible to talk about 'path creation' (Garud and Karnøe, 2001; Stack and Gartland, 2003, p. 489). Otherwise, it is argued, path dependence begins to sound like a form of historical determinism (Greener, 2002, p. 614).[5] Without doubt, such criticisms should be kept in mind while utilizing

the path-dependence approach in exploring the emergence of clusters, reminding us of the need to carefully examine and flesh out the 'contextual factors' as well as the 'processes' involved. To achieve this, we need, for instance, to seek answers to the following questions: What are the forces at operation that keep actors on a path? Exactly how do institutions and structures constrain actors on a path? What causes a closure of or some narrowing down in some previously feasible paths?

To provide satisfactory answers to such questions in turn requires detailed empirical work 'to gain additional insights concerning the means by which structures become so limiting, as well as the potential for breaking free from them' (Greener, 2002, p. 615), without forgetting that in path dependence the object of study is a 'history' and an attempt to fully grasp it requires 'history-friendly' theorizing (Feldman, 2001). How to conduct such empirical work poses challenges, and there are strong arguments favoring case studies in this regard. Specifically, apart from offering a detailed and holistic analysis of sequences in historical cases (Bennet and Elman, 2006, p. 250), using case studies enables analysts to clarify which parts of the account are contingent as well as the respective roles of each in subsequent events. Pre-existing conditions (contextual, institutional as well as social) in particular stand out among those that must be studied in detail. In short, it seems that a thorough understanding of the emergence of clusters requires the researcher to investigate the cluster's early development in detail, which favors and invites the use of historical methods (Clark and Rowlinson, 2004), and thus an identification of events, activities and choices over time (Jones, 2001, p. 918). This is what we aim to do in the following pages in order to explore the factors accounting for the emergence of Istanbul's film industry.

10.3 Emergence of the film cluster in Istanbul: why in Pera and why in Yesilcam within Pera?

Istanbul is a highly cosmopolitan city, geographically located at the juncture of the European and Asian continents. At the turn of the last century when cinema first appeared in Turkey, the financial center of Istanbul was the Galata district. Further up Galata lay the Pera district (today named as Beyoglu) where foreign embassies were – and, to some extent, still are – clustered, which was one of the main entertainment centers of the city.

By the end of the 1930s, film business in Istanbul had clustered considerably in the Pera district, especially on and around the district's

Grand Avenue, where offices of most of the film businesses came to be situated. This is probably not surprising since, in the words of Scognamillo (1991, p. 9), 'cinema chooses a city as its center in every country, be it a major city or the capital city. It develops and spreads by vitally connecting to that city. In Turkey, cinema has chosen Istanbul and an appropriate zone of that city, Beyoglu (Pera), as its center. This is neither an arbitrary nor a random choice.'

Unlike the case of the American film industry, whose westward migration meant a change of location from New York to Los Angeles, the city of Istanbul has been the host for the motion picture cluster throughout its history in Turkey. The reasons for Istanbul being the site of cinema's center in Turkey are indeed self-evident as it is the country's financial and cultural center. As we narrow down the geographical scale, however, the need to go beyond such general rationales becomes apparent. There were, for instance, without doubt candidates other than Pera (e.g., Sirkeci, a commercial center, or Sehzadebasi, the traditional entertainment center for the Muslim population) that might have hosted Istanbul's film cluster. Events, however, unfolded in such a way that Pera was, at the end, the location that began to show signs of endogenous development, and the reasons for 'cinema's choosing' Pera merits some further scrutiny.

The question in front of us is then the following: Why was it Pera rather than, say, Sirkeci which ended up generating a cluster dynamic? In fact, since we know that cinema had entered Istanbul via Pera – the first public screening taking place in a pub on its Grand Avenue in 1896 – we might even ask more specific questions than that: Which characteristics of Pera – and its Grand Avenue – 'invited' this debut to its site? And further, why did it remain clustered around Pera and especially its Grand Avenue despite the fact that cinema in these early years quickly spread to the other districts of the city?

In order to answer these questions, we first need to dig into the characteristics of Pera. Pera had initially emerged as the resort area of the city's *levanten* (merchants and financiers of Italian and other Western origins) community working in the banking district of Galata. In the nineteenth century, it had become a prominent shopping, leisure and entertainment locale, housing shops of luxury consumption goods, patisseries, casinos, cafes, stage theatres, a skating palace and a circus along its Grand Avenue, catering mostly to non-Muslim communities. Even exhibitions of pre-cinematic forms of moving images such as cosmoroma, magic lantern and diaphanorama had already been carried out sporadically in these locales (Evren, 2006, p. 15). Hence the facts that it was already an established entertainment center and that it was

populated by non-Muslim (local and foreign) communities obviously made it a natural, almost inevitable site for cinema, a novel entertainment mode originated in the West, to make its first appearance there. It can thus be argued that the main reason behind the very first screenings of cinematograph taking place in Pera was 'demand-related'; that is, induced by the presence of a 'sophisticated local demand' if we use Porter's (1998) terminology. The fact that the native minorities of the city acted as a connection point enabling and facilitating interaction with the inventors and developers of cinema in the West also contributed towards an increasing level of agglomeration of cinema in and around Pera, underscoring the parts played by 'historical accidents' (Krugman, 1991), 'specialized and advanced factor conditions' (Porter, 1998), and 'entrepreneurial careers' (Jones, 2001) in the emergence of the cluster.

However, it should be stated that after the first screenings in a pub on Pera's Grand Avenue, it would take only two months for screenings to begin in Sehzadebası in the Muslim sector. In a similar fashion, while the first permanent cinema hall was opened in Pera, cinema halls would appear all around the city within a few years. And yet, despite this proliferation, Pera would consolidate itself as a film cluster; in fact, as *the* cluster. It might be tempting to 'explain away' this fact through a cultural-essentialist reasoning that cinema was alien to the traditional Muslim Turkish domestic culture. This explanation, however, would not apparently be valid: on the contrary, as the first-ever historian of Turkish cinema, Rakım Calapala, had stressed, the Muslim masses had little difficulty in embracing cinema (Calapala, 1947, cited in Mutlu, 2002, p. 98); pointing to the fact that cafes showcasing traditional shadow plays were used to expose Turks to cinema, Arslan convincingly argues (2005, p. 31) that 'the import had become local' as an 'existing setup' was used 'to incorporate [it] into existing spectatorial practices.'

The true explanation as to why Pera consolidated as the cluster has more to do with the fact that cinema was a highly import-dependent venture at the time, and the first major film importers were from the non-Muslim communities, all clustered in the financial district of Galata, adjacent to Pera, and with well-established ties with the West. This community of early importers was situated at the heart of the city's banking cluster in Galata, adjacent to the clustering of foreign embassies (doubtlessly an important advantage in the import business) in Pera and to a dense concentration of some of the largest cinema halls on Pera's Grand Avenue. Consequently, it is not surprising to see that, in time, the offices of exhibition companies became conveniently clustered in Pera as well, the only exception being the Turkish-owned Kemal

Film exhibition chain based in Sirkeci, which, while being in the old city rather than the Galata-Pera region, was nevertheless not far away from it either. It is striking, however, to note that when another Turkish businessman, Halil Kamil, went into film ventures (as an exhibitor and importer), he felt it necessary to move his established offices from Sirkeci to Pera. Clearly, Pera was already gaining a 'magnetic' nature in respect to film business as early as the 1920s. Therefore, it is important to underline that when Turkish entrepreneurs decided to enter the industry, they were also being drawn to the 'center of gravity' in Pera even if they initially operated in the predominantly Muslim district of Sirkeci. The entry of more Turkish exhibitors into import ventures, most notably that of the Ipekçi Brothers, on the other hand, pushed the base of import business towards that center of gravity as well, which as a result began to shift to Pera while loosening up in Galata. Meanwhile, the same exhibition/import-based importers' entry into film production meant that the seeds of the production business also began to sprout in the Pera region. Finally, when distribution emerged as a distinct branch of cinema in Turkey, the first major distributor (Lale Film) would also open its offices in Pera.

During Turkish cinema's take-off years in the 1940s and 1950s – that is, when production of Turkish films picked up – not only the cluster in Beyoglu (Pera) further consolidated, but it condensed even more densely in one particular locale of Beyoglu, namely Yesilcam Street and its immediate environs (see Figure 10.1). Almost all of the major film companies established in the crucial 'sprouting' period 1944–1946, and several of the minor ones, set up their offices either on Yesilcam Street or Sakizagaci Street parallel to Yesilcam, or on Alyon Street opposite Yesilcam across Istiklal Avenue (Grand Avenue). What made Yesilcam Street, in particular, stand out among the many other streets along the avenue to attract all these new-comers calls for further elaboration.

Yesilcam, more than any other street along the avenue, was the site of an accumulation of several large cinema halls since the 1910s. The issue of how this street in particular – and not any other street along the avenue – had become the site of a dense cluster of cinema halls in the first place, can be tracked along its history. In the 1890s, the area next to one of the two blocks along Yesilcam (then called Deveaux Street) housed a horse circus (Cirque de Pera) and a variety theatre. The other block along Yesilcam housed the Odeon Theatre as well as one of Istanbul's first and largest apartment buildings, the Luxembourg Apartments, which served as a cafe and a hotel. The condensed presence of such popular sites appears to have been attractive to the first cinematograph

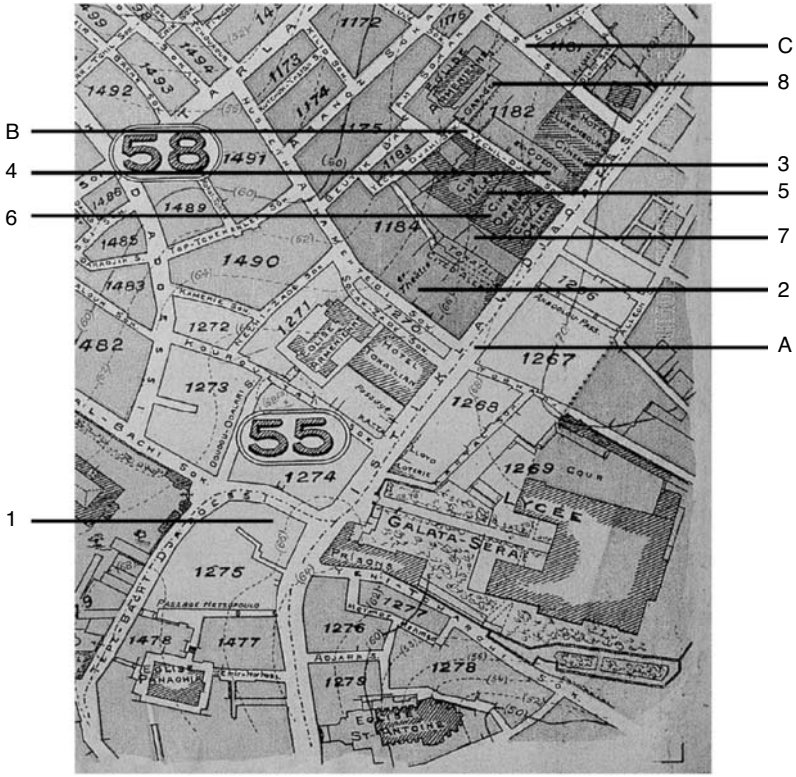


Figure 10.1 Map of the Yesilcam Street and its environs

Source: Vintage map drawn by Jacques Pervititch in 1928 (Anadol and Ersoy, 2001).

Notes:

A: Istiklal Avenue (Grand Avenue of Pera).
B: Yesilcam Street.
C: Sakizagaci Street.
1: Sponeck Pub – the site of first public screening in 1896.
2: Variety Theatre – site of one of the early public screenings circa 1897–1902; converted into Orientaux Cinema in 1911.
3: Luxembourg Hotel Cafe – site of one of the early public screenings circa 1897–1902; converted into the Gaumont Cinema in 1913 (later re-named Saray).
4: Odeon Theatre – site of one of the early public screenings circa 1897–1902; converted into the Eclair Cinema in 1914; run by Necip Erses (and re-named Sark) from 1933 onwards.
5: Skating Palace converted into the Ottoman Cinema in 1915, reopened as the Melek Cinema in 1924, run by the Ipekci Bros. and housing their offices.
6: The Opera Cinema opened in 1924; run by the Ipekci Bros. (and re-named Ipek) from 1933 onwards.
7: The Artistik Cinema opened in 1930 (later re-named Sümer).
8: The Ar Cinema, converted from garage, opened in 1943.

exhibitors as several of the very first cinematograph screenings in Pera are known to have been carried in this complex in the period circa 1897–1902: in the Variety Theatre, in the Odeon Theatre and in the cafe of the Luxembourg Apartments (Gökmen, 1989, pp. 13–14). In 1911, the Variety Theatre was converted into a permanent cinema hall, being the second of its kind in Pera and the first on Grand Avenue. Between 1913 and 1915, the Odeon Theatre, the Luxembourg Apartments, and a skating palace on the blocks along Yesilcam all followed suit; by 1943, three more cinema halls would be set up in these two blocks.

Hence by the early 1940s, the two blocks on each side of Yesilcam housed a total of six cinema halls, three on each block; most of these had more than 1,000 seats, making them among the largest cinema halls in the city. Two of these largest cinemas were run by the Ipekci Brothers who operated the largest chain of cinemas in the city, ran a film-import franchise with contractual ties to several US majors and were themselves one of the pioneers of film production in Turkey. Their offices were seated in one of these cinema halls in Yesilcam.

It is interesting to see that when we move from the level of the city and zoom into the level of the district and even to the level of the neighborhood formed around a single street, the true, underlying dynamics shaping a cluster's emergence and subsequent evolution are magnified. Thanks to the further concentration observed in Istanbul's film cluster, we are able to see, first of all, that a simultaneous presence of a 'sophisticated local demand' and 'a few successful companies' invited more activity of a similar kind to the area. Given that 'Turkish businessmen believe what they see' (Eraydin, 2002), they did not wait for long after seeing the success of others. The main symbols and/or forms (including other firms' location choice) are imitated immediately by other firms to evoke legitimacy (Schuman, 1995). The fact that the early managers of the first cinema halls prior to 1914 were all non-Muslims show that the city's minorities played important roles as a connection point (pointedly, one of the first movers to Yesilcam, the Ipekci Brothers, were from a Muslim-convert family of Jewish origin) in a business which was initially highly import dependant. Additionally, the fact that the founders of major companies had an exhibition/import background (see Table 10.1) reinforced the attractions of Pera – hosting many movie theatres (i.e. exhibition) and being near Galata (the then heart of the importing business) – as a location for the film industry.

When Yesilcam began to generate externalities, a self-reinforcing dynamic then attracted an even further concentration to the locale; this time also being associated with the presence of established networks,

Table 10.1 The early major Turkish film companies

Company	Location	Entrepreneur	Prior Career
Kemal Film	Sirkeci	Ali Efendi and Kemal Bey	Restaurant owner; film exhibition/import
Ha-Ka Film	Pera (moved from Sirkeci)	Halil Kamil Bey	Gramophone merchant; film exhibition/import
Ipek Film	Pera	Ipekci Brothers	Silk merchant; film exhibition/import
Lale Film	Pera	Cemil and Tevfik Filmer	Cameraman; film exhibition/import
Ses Film	Pera	Necip Erses	Film exhibition
Istanbul Film	Pera	Faruk Kenc	Film director
Halk Film	Pera	Fuat Rutkay	Insurance agent; film exhibition

Source: Compiled by the authors based on Gökmen 1989; Özgüc 1996.

information spillovers, specialized talent and a whole range of related and supporting industries. The magnetic nature of clustering, in other words, began to show its signs, turning Yesilcam into 'the place to be' if one wants to be in-the-know about this area of activity. All this, as a result, determined the shape of the path Istanbul's film cluster took, and thus how, if we borrow from Storper (1989, p. 300), 'history [was] produced' in this particular locale, locking this specific activity into Yesilcam and its immediate environs.

10.4 Concluding remarks: elementary, Watson!

We have set out to explore in this chapter the specifics of how and why a film cluster emerged in Istanbul's Yesilcam district, an endeavor, which is argued to be analogous to detective work (Bennet and Elman, 2006, p. 262), where the researcher identifies 'clues' and 'suspects' that might help explain the contingency, and examines the evidence in order to rule out some variables and to instantiate other variables or interactions as explanations of the contingent event. In our quest to find the reasons why the film industry is geographically concentrated in Istanbul, and in Pera within Istanbul, and finally in Yesilcam within Pera, we have uncovered that not only the events as they fall (Krugman, 1991) but also the attributes of the business environment that they fall into (Porter, 1998; Schuman, 1999) as well as the backgrounds of firms/entrepreneurs

(Jones, 2001) that reside within that environment seem to collectively determine the particular shape that a cluster's path takes.

As we have been warned (Martin and Sunley, 2006), while conducting this analysis, one should not become entrapped in a rather pointless search for one single 'ultimate cause' and/or over-emphasize the part played by 'historical accidents', which might result in placing too much emphasis on random events as the sources of path creation while neglecting to identify contextual and causal processes. It is not only that chance and skill are likely to be complements in that 'those initiatives that embody a superior business model or technology are more likely to find the "luck" they need' (Bresnahan et al., 2001, p. 842), leading us to question the degree to which chance is exogenous (if good entrepreneurs may create their own opportunity) (Feldman, 2001), but also that there might be severe constraints preventing entrepreneurs from successfully exploiting such 'accidents'. Take the case of the important part played by Istanbul's minorities in the emergence of the cluster as an example. An interpretation of this fact as a 'chance event' and 'the ultimate cause' explaining the emergence of the cluster would miss some key contextual circumstances such as the ways Muslim Turks embraced cinema and eventually became a part of the cluster, as discussed above. In fact, historical accidents might be happening all the time, and clarifying the circumstances under which they might successfully be exploited by the cluster participants seems to be the real challenge in front of us.

Without doubt, this does not rule out the possibility that sometimes mistakes and misperceptions of the entrepreneurs might have an effect on determining the direction of the path followed by a cluster, too, as theorized in the literature (Scott, 2005). Opportunistic behavior of the entrepreneurs in the 'golden age' of the Istanbul cluster (corresponding to a period from the late 1950s to the early 1970s), for instance, meant that enormous funds gained were not invested in infrastructure for the development of the industry, but, instead, used for shooting more and more films in an attempt to fully exploit the boom. The long-run consequences of this strategy were, however, severe: the path the cluster entered into was heading towards a cliff at full speed, which in turn reminds us that the transformation that a cluster goes through is not always progressive and that how the cluster participants respond to the challenges they face does matter. There is, in other words, a circular causality at work (Porter, 2006, p. 486) characterized by interdependence, forming the particular shape of the path that the cluster has been going through.

Another dimension of the issue is linked to the observation that, paralleling the case of the Paris cluster, Istanbul cluster's inner core was composed of central offices of the film companies and related businesses and movie theatres, whereas the studio lots themselves were concentrated around the Sisli/Mecidiyekoy districts, corresponding to the then outskirts of the city. Interestingly, we observe that when the Ipekci family took over the operation of two big movie theatres on Yesilcam Street and also decided to locate their offices in this area, Yesilcam Street gained additional dynamism in that the success of these ventures invited more investments of similar kind to their vicinity. In a similar vein, the decision by the same family to open up a studio in Nisantasi arguably had an impact on the subsequent location decision of Ha-Ka Co., which opened up a studio in nearby Sisli two years later. It indeed seems that cluster formation is a self-reinforcing process, and the location choice of additional investments/entries into the same industry tend to occur less and less randomly, as theorized by the path-dependence approach. Besides, newcomers are likely to bring their own contributions, and thus add their own touch to the formation of the specific path that the cluster would follow, as the entry of Istanbul Film Co., which introduced the 'star system', and Ses Film, which initiated the practice of shooting films without sound and dubbing them, illustrate.

The Istanbul cluster provides supporting evidence in favor of another observation as regards film clusters (Scott, 2002): the prominence of social networks. We know that Yesilcam hosted not only movie theatres and film companies but also cafes/restaurants/bars in which webs of relations and thus social networks were created and re-created. In relation to this, when we dig into the reasons why no fresh capital from other sectors flew into Yesilcam despite the fact that it was seen as a 'gold mine' for a significant period of time,[6] we can argue that the difficulty of entering into the 'close network' of relations might have played a role, as pointed out in the relevant stream of literature (e.g., Coe, 2000). One exception in this regard was an attempt by a well-established organization in Turkey (namely, Is Bank), which did try investing in this industry. The outcome of this venture, however, was a disappointment, and probably discouraged other entrepreneurs and/or organizations that had an appetite for the industry. This incidence shows, among other things, that information on failures – not only about successes – disseminates equally rapidly in a cluster context,[7] whose ultimate outcome might well become a lock-in effect to a particular path, instead of an otherwise likely scenario of swimming into other, perhaps brighter, waters.

Acknowledgements

An earlier version of this chapter was presented at the 23rd European Group for Organizational Studies (EGOS) Colloquium, Sub-Theme 14: Path Dependencies and Beyond (Vienna, 5–7 July 2007). We are indebted to the valuable comments made by the participants of the sub-theme, especially Peter Clark, Jörg Sydow and Philipp Tuertscher. Our thanks also go to Yeliz Günel for helping us with the map. Finally, Özlem Öz would like to thank The Turkish Academy of Sciences for their financial support.

Notes

1. An illustrative example comes from the Indian film industry, in that the particular path that Bollywood has been following cannot be fully understood without taking the Indian government's policies (e.g., those targeting a transition from black-market to legitimate financing of films in India, see Mehta, 2005, p. 139) into account.
2. Empirical evidence suggests that there are examples of film clusters where national and transnational organisations have proved more important than localised networks (e.g., internal linkages within transnational corporations may substitute for cluster linkages) (Turok, 2003). Others (e.g., Enright, 2000, p. 115) emphasize an interdependent model 'in which foreign multinationals play a critical role in the cluster, and location in the cluster plays a critical role in the strategy of the foreign multinational'.
3. The principle changes that have had particularly strong impacts in this regard include wars, major changes in rules and regulations, the advent of TV and more recently of DVD and new computerised technologies (Scott, 2002). Misery that the war (WWII) years brought, for instance, ironically triggered an improvement in the fortunes of Hollywood, as it generated additional motivation for an escape into fantasy.
4. Our interpretation of lock-in is close to Callon's (1991) concept of irreversibility, being defined to be dependent upon two factors: 'first, the extent to which we can return to a point where the dominant translation is just one amongst many; second, the extent to which the translation which is presently dominant shapes and determines future strategies' (Greener, 2002, p. 617).
5. Concepts borrowed from logic – namely, sufficiency and necessity – may help clarify this point: sufficiency implies that once the first step on the path is taken the final outcome is very likely to happen, whereas necessity implies that there may be conditions without which the next step in the chain would not have been possible (Bennet and Elman, 2006, p. 258).
6. By way of comparison, note that the world's leading film cluster in Hollywood had considerable capital inflow from a variety of sources via Wall Street, see Cook and Bernink, 2002.
7. In fact, on some occasions even more rapidly than success stories, see Öz, 2004.

References

Aharonson, B.S., J.A.C. Baum and M.P. Feldman (2007) 'Desperately seeking spillovers? Increasing returns, industrial organization and the location of new entrants in geographic and technological space', *Industrial and Corporate Change*, vol. 16, pp. 89–130.

Aldrich, H.E. (1999) *Organizations Evolving*. Thousand Oaks: Sage.

Anadol, C. and S. Ersoy (eds) (2001) *Sigorta Haritalarinda Istanbul*. Istanbul: Tarih Vakfi and Axa Oyak.

Arslan, S. (2005) *Hollywood Alla Turca: A History of Popular Cinema in Turkey*, Unpublished PhD Dissertation, Ohio State University, USA.

Arthur, W.B. (1985) 'Competing technologies and lock-in by historical events: The dynamics of choice under increasing returns', *CEPR Technical Paper No. 43*. Stanford: Stanford University.

Bennet, A. and C. Elman (2006) 'Complex causal relations and case study methods: The example of path dependence', *Political Analysis*, vol. 14, pp. 250–267.

Bresnahan, T., A. Gambardella and A. Saxenian (2001) 'Old economy inputs for new economy outcomes: Cluster formation in the new silicon valley', *Industrial and Corporate Change*, vol. 10, pp. 835–852.

Brusco, S. (1996) 'Global systems and local systems', in F. Cossentino, F. Pyke and W. Sengenberger (eds) *Local and Regional Response to Global Pressure: The Case of Italy and Its Industrial Districts*. Geneva: ILO, pp. 145–158.

Callon, M. (1991) 'Techno-economic networks and irreversibility', in J. Law (ed.) *A Sociology of Monsters: Essays on Power, Technology and Domination*. London: Routledge, pp. 132–165.

Clark, P. and M. Rowlinson (2004) 'The treatment of history in organisation studies: Towards an historic turn?' *Business History*, vol. 46, pp. 331–352.

Coe, N.M. (2000) 'The view from out West: Embeddedness, inter-personal relations and the development of an indigenous film industry in Vancouver', *GeoForum*, vol. 31, pp. 391–407.

Cook, P. and M. Bernink (2002) *The Cinema Book*. London: BFI.

David, P.A. (1985) 'Clio and the economics of QWERTY', *American Economic Review*, vol. 75, pp. 332–337.

Dudrah, R.K. (2006) *Bollywood: Sociology Goes to the Movies*. New Delhi: Sage.

Enright, M.J. (2000) 'Regional clusters and multinational enterprises: Independence, dependence, or interdependence?' *International Studies of Management & Organization*, vol. 30, pp. 114–138.

Enright, M. (1990) *Geographical Concentration and Industrial Organization*, Unpublished PhD Dissertation, Boston, Harvard University.

Eraydın, A. (2002) 'The local embeddedness of firms in Turkish industrial districts: the changing roles of networks in local development', in M. Taylor and S. Leonard (eds) *Social Capital and the Embedded Enterprise: International Perspectives*. Aldershot: Ashgate, pp. 269–288.

Evren, B. (2006) *Türk Sinemasi / Turkish Cinema* (Istanbul-Antalya: Aksav-Türsak).

Feldman, M.P. (2001) 'The entrepreneurial event revisited: firm formation in a regional context', *Industrial and Corporate Change*, vol. 10, pp. 861–891.

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

Gökmen, M. (1989) *Baslangıctan 1950'ye Kadar Türk Sinema Tarihi ve Eski Istanbul Sinemalari*. Denetim Ajans Basimevi, Istanbul.
Greener, I. (2002) 'Theorizing path-dependence: How does history come to matter in organizations?' *Management Decisions*, vol. 40, pp. 614–619.
Jones, C. (2001) 'Co-evolution of entrepreneurial careers, institutional rules and competitive dynamics in American film, 1895–1920', *Organization Studies*, vol. 22, pp. 911–944.
Krugman, P. (1991) *Geography and Trade*. Cambridge, MA: The MIT Press.
Lewin, A.Y. and H. Volberda (1999) 'Prolegomean on coevolution: A framework for research on strategy and new organisational forms', *Organisation Science*, vol. 10, pp. 519–534.
Lukinbeal, C. (2002) 'Teaching historical geographies of American film production', *The Journal of Geography*, vol. 101, pp. 250–260.
Markusen, A. (1996) 'Sticky places in slippery space: A typology of industrial districts', *Economic Geography*, vol. 72, pp. 293–313.
Marshall, A. (1949) *Principles of Economics: An Introductory Volume*. London: MacMillan.
Martin, R. and P. Sunley (2006) 'Path dependence and regional economic evolution', *Journal of Economic Geography*, vol. 6, pp. 395–437.
Martin, R. and P. Sunley (2003) 'Deconstructing clusters: Chaotic concept or panacea?' *Journal of Economic Geography*, vol. 3, pp. 5–35.
Mehta, M. (2005) 'Globalizing Bombay cinema: Reproducing the Indian state and family', *Cultural Dynamics*, vol. 17, pp. 135–154.
Mutlu, D.K. (2002) *Yesilcam in Letters: A 'Cinema Event' in 1960s Turkey from the Perspective of an Audience Discourse*, Unpublished PhD Dissertation, Bilkent University, Turkey.
Öz, Ö. (2004) *Clusters and Competitive Advantage: The Turkish Experience*. London: Palgrave MacMillan.
Özgüc, A. (1996) *Türk Film Yapimcilari Sözlügü*. Istanbul: Fiyap.
Porter, T.B. (2006) 'Coevolution as a research framework for organizations and the natural environment', *Organization & Environment*, vol. 19, pp. 479–504.
Porter, M.E. (1998) *On Competition*. Boston: Harvard Business School Publishing.
Porter, M.E. (1990) *The Competitive Advantage of Nations*. New York: The Free Press.
Schuman, M.C. (1995) 'Managing legitimacy: Strategic and institutional approaches', *Academy of Management Review*, vol. 20, pp. 571–610.
Scognamillo, G. (1991) *Cadde-i Kebir'de Sinema*. Istanbul: Metis Yayinlari.
Scott, A.J. (2005) *On Hollywood: The Place, the Industry*. Princeton and Oxford: Princeton University Press.
Scott, A.J. (2002) 'A new map of Hollywood: The production and distribution of American motion pictures', *Regional Studies*, vol. 36, pp. 957–975.
Scott, A.J. (2000) 'French cinema: Economy, policy and place in the making of a cultural-products industry', *Theory, Culture & Society*, vol. 17, pp. 1–38.
Stack, M. and M.P. Gartland (2003) 'Path creation, path dependence, and alternative theories of the firm', *Journal of Economic Issues*, vol. 37, pp. 487–494.
Storper, M. (1999) 'The resurgence of regional economics, ten years later: The region as a nexus of untraded interdependencies', in J. Bryson, N. Henry,

D. Keeble and R. Martin (eds) *The Economic Geography Reader: Producing and Consuming Global Capitalism*. Chichester: Wiley, pp. 209–215.

Storper, M. (1989) 'The transition to flexible specialization in the US film industry: The division of labour, external economies and the crossing of industrial divides', *Cambridge Journal of Economics*, vol. 13, pp. 273–305.

Storper, M. and A.J. Venables (2004) 'Buzz: face-to-face contact and the urban economy', *Journal of Economic Geography*, vol. 4, pp. 351–370.

Sydow, J., F. Lerch and U. Staber (2010) 'Planning for path dependence? The case of a network in the Berlin-Brandenburg optics cluster', *Economic Geography*, vol. 86 (in print).

Turok, I. (2003) 'Cities, clusters and creative industries: The case of film and television in Scotland', *European Planning Studies*, vol. 11, pp. 549–565.

11

Path Dependence in Regional Development: What Future for Baden-Württemberg?

Gerhard Fuchs

11.1 Introduction

The extent to which regions can diverge from established paths of economic development continues to be a matter of heated debate. There are potent arguments that any new constitution of regional economies is determined to a considerable degree by institutional and industrial structures that have emerged in the course of regional industrialization history, often spanning hundreds of years. Such structural lineage is often considered to hinder the adjustment to new challenges. The increasing globalization of national and regional economies, the rise of supra-national economic blocs, ongoing developments in technology, and new organizational strategies are among the elements that individually and collectively add new spins to this path dependence debate. Simultaneously, these challenges bring along new possibilities of innovation, and the transfer and adaptation of knowledge. New technologies offer new perspectives and institutional structures; learning systems and policy mechanisms might serve as change factors that promise fresh opportunities for regions to pursue new development trajectories.

In this chapter we seek to cast light on both the importance of economic-structural and institutional path dependencies. Undoubtedly there are both technological (cf. Dosi, 1982) and regional trajectories. Technological knowledge is not only organized in large-scale technical systems (Hughes, 1987), in industries or in professions but frequently also in regional economic areas, where it is embedded in a broad institutional and cultural context. Knowledge is also incorporated in regional production clusters, cooperative relations, institutions, and policy patterns

where it does not usually develop in great leaps and bounds, but incrementally, in a mostly evolutionary manner.

Given today's global economic conditions and frequently expressed notion of the 'knowledge society', the development and institutionalization of collaborative mechanisms of learning is assuming increasing importance – elevating the importance and role of collective levels of reflection. It has been observed that 'technological capabilities [...] reflect local, regional and national contexts and environments' (Storper, 1995, p. 897). However, while we assume that technical knowledge and technological learning are influenced by context and region, this still raises issues about the extent to which learning processes track a path-dependent course and, if so, exactly how that learning path correlates with and is bounded by the development paths of technologies and regions.

One option to increase regional innovative capacity is by strengthening, re-orientating, or creating institutions – fostering, in effect, what Ash Amin and Nigel Thrift (1994) have termed 'institutional thickness.' They remind us that this concept has multiple implications. First, the staying power of institutions is of considerable significance in regional development. Second, the local construction and enrichment of a reservoir of shared knowledge is a notable element of regional dynamics. Third, the capacity for learning and change is inherent to institutional flexibility within regions. Fourth, the innovative capacity of companies is viewed as a shared characteristic of a given region. Fifth, regional interactions are firmly based on trust and reciprocity. Finally, regions give rise to a consolidated feeling of belonging among their inhabitants, reinforcing regional social capital. For Amin and Thrift, regional trajectories are secured partly by established economic structures and production clusters, and partly by regional institutions. Particularly in innovation research, which tends to concentrate on industrial and technological changes, the importance of the region as a collective sphere of economic activities, institutional and social relationships, and political negotiations cannot be underestimated.

This line of argumentation suggests that the economic position of a region within the context of global competition is partly the result of path-dependent developments, but is also influenced by current institutionally anchored governance structures that influence the regional innovation capacity. The institutions of the regional innovation system not only serve as a resource for firms to draw on (technological know-how, qualified workforce, cooperative labor relations, etc.), but also fulfill important orientation and governance functions for regional actors

in industry, science and politics. As institutional thickness increases and gives rise to institutional inertia, the industrial development paths of a region also become institutionally stabilized; technological path dependencies are thus accompanied by institutional ones. Given this, we have to explore how to deal with the challenges involved in reforming and developing regions and their industrial, innovation, learning, and institutional systems. Within the context of globalization, national and regional actors in industry and politics are faced with the challenge of constantly reappraising the functional and operational principles of regional innovation systems and, if necessary, devising strategies for reforming the institutional and industrial order (cf. Cooke, 1998; Drache, 1996).

This chapter will analyze the development of the regional innovation system of Stuttgart/Mittlerer Neckar (the core region of Baden-Württemberg, Germany) which at the same time exhibits strong features of successful path-dependent development as well as clear indications for a strong culture of learning and adaptation when faced with new challenges. We analyze a case of continuous institutional adaptation and development, which allows us to say that in spite of severe changes within the entity under consideration, there is still 'a' path and a quasi entity creating it while using it. An important role in adjusting the adaptive capacities of the industrial structure is played by institutional developments. These institutional developments are aimed at avoiding lock-in effects. Institutional developments have not changed the salience of the industrial structure of Stuttgart/Mittlerer Neckar, but have increased and strengthened the adaptive capabilities available in the region. As such, the chapter will highlight institutional developments and variation over the last two decades, employing Kathleen Thelen's (2004) concepts of institutional layering and institutional conversion for the analysis of path-dependent processes. The study is based on and summarizes various empirical studies on the regional innovation system of Baden-Württemberg that have been conducted by the author over the last ten years. In order to make our point we will start (Section 11.2) with a summary of the literature claiming that Baden-Württemberg has been a victim of path-dependent developments resulting in a severe economic crisis in the early nineties and a discussion of potential lock-in factors (Section 11.3). The next part will highlight that the present success of the region can, however, also be attributed to the region following its established path of industrial strengths (Section 11.4). Section 11.5 will discuss institutional developments which aimed at developing a new path for the region in the

1990s, failed to do his, but in the end (unintentionally) secured that the established path could be successfully followed (Section 11.5), before finally drawing some conclusions (Section 11.6).

11.2 The issue: continuing positive feedbacks or stuck in a dead-end street?

In the early-to-mid-1990s various scientific and non-scientific pundits predicted a dire economic future for the region of Baden-Württemberg – until then an economically outstandingly successful 'Bundesland' in the south-west of Germany. The limitations of Baden-Württemberg's regional production system seemed to become increasingly apparent. During the economic crisis between 1992 and 1994, the economy of Baden-Württemberg suffered more than the other states of former West Germany. The GDP declined by 4.7 per cent (1993), while unemployment rate rose from 3.7 per cent (1991) to 8.7 per cent (1997). This crisis primarily affected the economic core region of Baden-Württemberg, the region of Stuttgart/Mittlerer Neckar (see Krauss, 1999: 359). The automotive, mechanical engineering, and electronics and electrical engineering industries all underwent processes of personnel retrenchment. Between 1980 and 1995, more than 15 per cent of the employment in mechanical engineering was lost; in the electronics and electrical engineering industry, the decline amounted to almost 14 per cent of the jobs. Employment related to the production of electronic data-processing systems declined by 10.5 per cent between 1980 and 1996 (see Statistisches Landesamt, 1997). The central thesis voiced in these years was that Baden-Württemberg was affected by a process of path-dependent development which, although very successful in the past, would lead the region's economy into a potential dead-end street. It was expected that that by 2010 the old dominating industries would have collapsed. In 1994, one regional research institute cautiously predicted employment in Baden-Württemberg to drop even further until 2010: in the electronics and electrical engineering industry by 9 per cent, in the mechanical engineering industry by 15 per cent and in the automotive industry by 26 per cent (Saebetzki, 1994, p. 92).

A break away from the established paths of economic development seemed to be necessary in order to guarantee future sustainable economic growth. Now (in 2009) we should be able to ask ourselves whether this analysis based on thinking in path-dependent terms was correct. Has Baden-Württemberg in fact lost its leading position due to its concentration on established strengths or have significant changes taken

place that have altered the contours of the Baden-Württemberg model and made it more adaptive to changes in the world economy?

Path-dependent phenomena are caused by positive feedbacks. This implies that history matters in the sense that historical paths once chosen, even by coincidence, then rigidify and dominate future developments. A famous example of technological path dependence is the often-quoted QWERTY phenomenon of the typewriter keyboard. The typewriter keyboard arrangement solved a temporary mechanical problem on one of the first typewriters. But it soon became the standard for generations to come, even though it has been considered by many to be inefficient and other, potentially better, solutions were available (David, 1985; Arthur, 1994).

Technological skills reflect local, regional, and national contexts and environments. Regions that are extraordinarily successful are those where the institutional context and regional networks closely complement the dominant industrial clusters of the region. Regional paths of development thus refer to accumulated competencies, methods, and technologies. This regional stock of knowledge is further developed within regional networks and can be closed towards inputs from outside. These processes of closure, however, might lead to the development of lock-in situations that prevent regional institutions from adapting to changes in the environment and from learning new knowledge.

Path dependence can be maintained or fractured in particular ways in different locations. We will differentiate between three types: lock-in, incremental change, and breakthrough. Breakthrough means revolutionary changes of crucial economic structures and related institutions. Incremental changes mean evolutionary changes that do not set up new paths but try to lead old paths into a new direction (see Ackermann, 2001). Lock-ins, in contrast, are characterized by institutional and technological structures which are inefficient but cannot be changed. Grabher (1993) differentiates between three kinds of lock-in: functional, cognitive, and political lock-in.

Functional lock-in means that close cooperation within strongly tied networks impedes contacts with other regions. Future trends that occur outside these networks may be overlooked, with external resources poorly identified and used. Cognitive lock-in means that personal relations lead to shared common ideas, feelings and beliefs that prevent the adoption of new ideas. Political lock-in describes situations where historical trajectories of economic development are maintained by cooperative relations between regional actors which are then unable to see and adopt new ways of thinking and policy-making. Generally, it

can be assumed that breakthroughs are difficult to achieve in the case of existing institutional thickness and established, distinct economic structures within a certain region.

11.3 The main pillars of regional success

Baden-Württemberg has often been called one of the 'Four Motors of Europe' – a group of European regions regarded as being particularly successful in generating economic growth. It has earned this 'nomination' thanks to its outstanding economic success. Baden-Württemberg's research intensity is way above average. Its employment rate is also well above the German and European level; its unemployment rate significantly below these levels and its rate of European patent applications per million people is the highest of all European NUTS2 regions. The federal state of Baden-Württemberg, in the southwest of Germany, has thus been considered by many researchers to be a regional model economy up to the early nineties (Cooke and Morgan, 1990; Gabriel, 1990; Hassink, 1992; Maier, 1989; Sabel, 1989; Schmitz, 1992; Semlinger, 1993). With high rates of industrial investment and export, a reputation for high quality, and well-engineered products, it seemed to have overcome the major problems of many regional or national economies: namely how to establish and maintain competitive advantage in spite of severe structural changes in the economy. A significant number of specialized and export-oriented capital goods producers, a dominance of flexible small and medium-sized enterprises (SMEs), and a high capacity for technological innovation within these firms were seen as decisive characteristics. Its base of skilled labor, cooperative industrial relations, well-developed research infrastructures, state and national industrial policies, and close and long-term relations between banks and companies were named as important ingredients of the success story. Social scientists interested in explaining the region's economic success laid emphasis on these institutional factors (Wallace, 1994, p. 68). They characterized the region as coming close to the model of an industrial district (see Sabel, 1989; Schmitz, 1992; Semlinger, 1994; Rehfeld, 1995) and being endowed with a relatively self-contained regional economy and system of governance (Amin and Thrift, 1994, p. 7). In the same way Cooke (1998), in his typology of Regional Innovation Systems, considers Baden-Württemberg to be a role model of a highly networked system, with a high degree of coordination between connected actors (e.g., companies, banks, research institutions, etc.).

As mentioned before, economic problems began to strike this model region in the early 1990s. Drawing on the general discussion about economic development and industrial policy in Baden-Württemberg, critics held that Baden-Württemberg's past success might have become the very reason for the region's comparatively slow orientation toward new industries. Grabher (1993) has aptly described how the Ruhr region became locked into a once successful path of development, leading the region into a deep economic crisis. According to Morgan (1994, p. 11), it would be 'surprising' if not at least some of the problems described by Grabher were applicable to the Baden-Württemberg case. Indeed, Braczyk et al. (1995) identify several of these problems. They point to the slow adaptation of Baden-Württemberg's industry to the 'Japanese challenge' as a case of cognitive lock-in and to the bias of Baden-Württemberg's technology policy towards the core industries as a case of political lock-in. They also detect a functional lock-in in Baden-Württemberg's high level of economic integration: the fact that the three industrial core sectors are so closely interwoven makes the whole region particularly vulnerable to economic crises.

In a similar vein, Heidenreich and Krauss (1998, p. 223) point out that the major companies in Baden-Württemberg's core sectors are quite reluctant to source services from external providers. As a result, opportunities for communication and cooperation outside the established trajectories are missed and economic restructuring is hampered by barriers to learning. Heidenreich and Krauss (1998, p. 229; see also ifo, 1995) furthermore showed that Baden-Württemberg's system of R&D and technology transfer strongly concentrates on the three industrial core sectors and on only advanced technology rather than on the most advanced areas of technology.

Given Stuttgart's exceptional success in capital goods industries, the question was raised why its economy did not succeed in staging a stronger reorientation to sectors promising greater growth potential. It seemed that Stuttgart's economy was failing to stake out greater shares in new areas such as information and communications technology, new materials, biotechnology, environmental and power technology, micro-systems technology, and production-oriented services (Faust et al., 1995), although the industrial, structural, and institutional preconditions in each of the cited areas were certainly not unfavorable. These questions lead invariably to the issue of the inertia of established production structures.

The assumption behind the questions asked by critics of the regional innovation system of Stuttgart is obvious. Innovativeness is considered to be the result of communication and cooperation opportunities

outside of historically evolved and institutionally and organizationally reinforced trajectories. Heidenreich and Krauss (1998) assumed that such communication and cooperation opportunities did not sufficiently exist in Stuttgart. The implication was that barriers to learning and not the maturity of Stuttgart's product range formed a major obstacle on the road to new innovation-promoting company strategies.

11.4 Still on the path in 2009?

During the second half of the 90s, and especially in the first years of the new century, the region experienced a remarkable revival; it became a leader once again. Was this the result of leaving the established path behind, creating a new one, was it still business as usual but in a changed environment, or were gradual reforms responsible?

First one has to clearly state that important elements that characterized Stuttgart/Mittlerer Neckar are still determinant. In comparison to Germany (and the EU), the Baden-Württemberg regional R&D system continues to be characterized by a specialization on technologies relevant for the construction of machines and cars. The dominance of the manufacturing sector has not been severely challenged. In Stuttgart/Mittlerer Neckar, three industrial sectors still remain particularly important measured in terms of both employment and sales. Among these three, the car industry stands out more than ever before. Stuttgart maintains and, indeed, has further developed 'the largest, thickest, and the most powerful auto cluster in Europe' (Morgan, 1994, p. 37). As this cluster has strongly influenced not only the region's economy but also its institutions and culture, it is no exaggeration to portray Stuttgart still as a 'car city' (see also Kaiser, 2007). The global automotive sector underwent a severe process of restructuring, which had and has important impacts on primary manufacturers as well as their suppliers. This reorganization is taking place on a worldwide scale and along the entire value chain. This restructuring, however, has helped Stuttgart manufacturers to stay on top and strengthen their position.

There are approximately 100 companies in Stuttgart's automotive sector. They are responsible for 43.6 per cent of the industrial output in the region. This signifies an important growth trend. In 1980, car manufacturing accounted for 28.7 per cent of the total output. In 1992, its share had dwindled to 27.3 per cent, but since then has steadily risen to close to 50.0 per cent in 2007. 25.9 per cent of all industrial employees are to be found there, with employment growing faster than in the rest of Baden-Württemberg as well as in Germany.

This development is in some way special to the Stuttgart region. Not all car manufacturers in Germany and especially the core manufacturing regions were as 'lucky' as Stuttgart with Daimler, Porsche, and Bosch (the world's second largest automotive supplier). Stuttgart did better than the rest of Baden-Württemberg as well as the rest of Germany. A special feature is again the export orientation. In 2007, exports amounted to 67 per cent of all production, up from 57 per cent in 1996. Car manufacturing is also very research-intensive. Close to 70 per cent of the companies there spend more than 3.5 per cent of their turnover on R&D. All the important functions for car manufacturing are present in the region.

Stuttgart is also a leading region in Germany with respect to patent activities. This holds both true for the number of patents per 100,000 inhabitants as well as employees (see Koschatzky and Jappe, 2004). Patent statistics demonstrate clearly that the Stuttgart economy develops along stable trajectories. Patents are concentrated in car manufacturing and mechanical engineering, with a somewhat lower specialization in electrical engineering. The data furthermore show that over the last decade new patent specializations have not been established. The degree of specialization has in fact even increased for the case of car manufacturing and to a more limited degree in mechanical engineering. The most important single companies with respect to patenting activities remained Bosch, Porsche, Daimler, and Alcatel (although the latter with much reduced significance). The technological strengths of the region are undoubtedly with car manufacturing and technologies of special relevance for car manufacturing, including car-related IT. The combination of excellence in these fields remains an essential comparative advantage. Compared to the recession year of 1993 in which every 17th job was dependent on the automobile industry, this relationship has jumped today to every 14th job. Considering only the manufacturing sector, dependence on automobiles has increased from every sixth to every fifth working place. Between 1995 and 2000, more than 50 per cent of the increase in the region's valued-added output can be contributed to the car industry, and its share of regional industrial output has increased from 16 to 21 per cent.

In summary we can conclude that industrial sectors that were identified as being in a 'critical' condition in the 1990s continue to dominate the Stuttgart regional economy. This also holds true for the aspects characterized as being 'critical' by the observers in the nineties:

- The share of the Baden-Württemberg economy in the worldwide trade of R&D-intensive goods continues to decline.

- The regional economy has still not secured a strong position in advanced services. Of course, there has been growth in services, but this increase mostly reflects the general German trend. Data do not reveal that in Stuttgart, or in Baden-Württemberg for that matter, services have any special, pronounced growth dynamic.
- The regional economy is still not quicker in adopting organizational reforms than the rest of Germany – a feature which was characterized as an instance of cognitive lock-in.
- The interlinkages between the main industrial sectors (functional lock-in) are still significant. A downturn in the automotive sectors directly influences the regional electronics and machine tool industry.

Overall, this presents us with a conflicting picture: On the one hand, we see that most of the traits of Stuttgart's economy that have been considered to be problematic are still there, in certain ways even extenuated. But we also see that the Stuttgart region remains one of the leading regions of Germany and has done better in recent years than the rest of the (less industrialized and more service-oriented) country. The strong concentration on specific technological fields was a major reason for the above average performance. In this respect, the crisis interpretation of the mid-1990s has to be reassessed.

11.5 Institutional re-alignments

Institutions can change due to endogenous processes, exogenous shocks, and combinations of both. The exact mechanism that brings about institutional change once the behavior associated with an institution is no longer self-enforcing depends on the nature of the quasi-parameters that delimit self-reinforcement. If these quasi-parameters are observable and their importance well understood, decision makers mighty actually realize that past behavior is no longer self-enforcing and the mechanism directly leading to institutional change will be intentional.

At an early stage, the state government of Baden-Württemberg was aware of a potential lock-in situation in the region's economic development. The person symbolizing the search for a new role for Baden-Württemberg and its core region of Stuttgart was the then federal Prime Minister Lothar Späth. Already in the late 1980s, he tasked expert commissions to consider the situation and make recommendations. This resulted in a number of activities to tackle lock-in phenomena and to develop new industry and service sectors and led to the creation of a number of new organizations in the field of science. Most prominently

among the sectors targeted were biotechnology, multimedia, and producer-oriented services. Späth was aware of the potential difficulties, but reinterpreted these as opportunities worth taking a risk for. In this sense, he viewed globalization as an option to be taken up proactively. In the industrial sector, Jürgen Schrempp – the most outspoken representative of treating globalization as a positive option – became the head of the later DaimlerChrysler. While the activities of the politicians were mainly in the realm of image building, changing the priorities of funding, looking for new policy instruments, and constructing new organizations, Schrempp transformed the regionally based company Daimler-Benz into the globally oriented and globally based company DaimlerChrysler. Let us look at some of these developments in more detail.

One means of increasing regional innovative capacity is by creating institutions. This often proves an inadequate solution, however (see Amin and Thrift, 1994). Harder to achieve is the development of a new regional identity and the generation of synergy effects between institutional and technological development paths. This is the task faced by especially those regions which have been very successful in the past, and which have achieved a high level of technological competence in so-called mature industrial sectors. For these regions, the hitherto established institutions and the institutional thickness thus achieved can even become an additional problem, since training, research and funding facilities tend to stabilize the traditional patterns of industrial development. Against a background of intensified global competition, this problem of institutional inertia and restrictions deserves special attention.

Since the late 1980s, the state government of Baden-Württemberg has been fully aware of the problems of institutional inertia. The response of the state government was a strategy that combined parallel institution building with a reform of old institutions (conversion). Parallel institution-building means that new organizations were created for the support and promotion of new industries. For instance, in the multimedia sector the Medien-und Filmgesellschaft Baden-Württemberg (Media and Movie Association of Baden-Württemberg) was founded. In the case of biotechnology, a state agency was established and various regional forms of institution building have been supported. It is noteworthy that these attempts – at least in the beginning – totally bypassed the established institutional structures and had few connections with these structures.

From the regional perspective, the most important element has been the foundation of the Association Region of Stuttgart in 1994. By this

a new regional identity was supposed to be developed. No longer was the city of Stuttgart and its identification as a 'car city' with strong local cultural roots the point of orientation, but a new region of Stuttgart, which pictures itself as a multimedia region, a hip-hop city, a biotechnology cluster and many other things, but determinedly trying to avoid the label 'car.' This process was initiated in the form of a decision by the state government. No other region became equipped in a similar manner. The most important task of the association became the furthering of economic development. It was clear from the beginning that this association should be primarily concerned with new sectors and new types of activities to support structural change in the region. The association is the product of a top-down development, represents the core of a new regional order, and is in the center of other institutional reforms as well. The association was active in the development and implementation of major projects. It organized regional capacities and resources in order for the region to be able to participate in national as well as European competitions for funding. Complex partnerships were built up, organized, and moderated.

The operationalization of the aims of the association was done gradually via concrete projects. In an exemplary manner one can name projects for site development, attracting new investors, network projects like the Regional Dialogue Car Production, MediaRegio Stuttgart, PUSH!, Mobilist, BioRegio, regional competence centers, the support program environmental technology and a host of marketing and image campaigns. Central fields of action are the promotion of innovations (not technologies) and the support of cooperation (see Table 11.1). On the one hand, this is the support for the development and use of new products and production processes; on the other hand, it is the instigation and support of cooperation in order to make use of potential synergetic effects in the region. In this sense, the association pursues a structural policy strategy, aiming at the support and stabilization of regional clusters, which should be accomplished by a network and innovation-oriented regional policy.

Table 11.1 New fields of activities in the region of Stuttgart

Cluster policy	**Employment promotion**
Promotion of Innovations	Regional reporting on important economic trends
Marketing/image campaigns	Regional competitions

The aforementioned attempts all aimed at changing the institutional structures by adding new layers. An example for institutional conversion is the reform of the system of universities of applied sciences. This type of university was originally thought to fulfill the demands of the local industry. The severe economic crisis in the first half of the nineties and a temporally rapidly decreasing demand for university graduates by regionally based companies, led to a dramatic situation (e.g., few new students) in many universities of applied sciences. The answer was twofold: on the one hand, traditional curricula were dramatically reduced; on the other hand, new curricula were developed, which aimed at decreasing the dependence on specific regional or local demand structures. Even smaller universities were now faced with the demand to develop 'world class' specializations. The regional orientation thus has been reduced and the importance of the openness to supra-regional developments and actors strengthened.

11.6 Conclusion

As remarked by Pierson and Skocpol (2002, p. 698), studying institutions usually means to analyze processes over a substantial stretch of years, maybe even decades or centuries. Our study was concentrated on a relatively short period of time, but took into account that the situation in the early 90s has been the product of a longer historical process. We hold that institutions can change due to endogenous processes, exogenous shocks, and combinations of both. In the case of the region of Stuttgart/Mittlerer Neckar it was both the external shock in the form of an economic crisis in the first half of the nineties, but also preparatory work already initiated beforehand in order to steer the region along a new path into different areas. Institutional survival in our case involved 'significant political renegotiation and heavy doses of institutional adaptation, in order to bring institutions inherited from the past into line with changes in the social and political context' (Thelen 2004, p. 8). The paper stressed that purposeful human action as well as the structural preconditions and contextual influences shaped the specific Baden-Württemberg path. Structurally the region of Stuttgart remains dominated by its industrial core; institutionally the region has opted for diversity and building global linkages, thus strengthening its learning capacities.

Grabher (1993) differentiated between three kinds of lock-ins that were considered to be also at work in the Stuttgart region: functional, cognitive, and political lock-in. As previously discussed, functional

lock-in implies that close cooperation within strongly tied networks impedes contacts with other regions: future trends that occur outside are overlooked and external resources are poorly identified and used. With respect to functional lock-in, we have analyzed the situation especially for the case of the automotive sector, which seemed to have successfully further developed its strategic position by developing itself into a truly global player and extending regional networks into a global arena.

Functional lock-ins are intensified through cognitive lock-ins. Personal relations lead to shared common ideas, feelings and beliefs. The fact that entrepreneurs in Stuttgart did not notice Japanese competitors at an early stage and that they neglected to introduce countermeasures in time, was interpreted as a cognitive lock-in. Collective misinterpretations were the reasons why Stuttgart did not evaluate its industrial structure in the context of global division of labor as well as division of technological competences. With respect to cognitive lock-ins it can also be asserted that starting in the mid-90s, Stuttgart began to reverse its image of being an old industry region. All major actors were looking towards new possibilities and options. Attempts to build up competencies beyond the core industries (as described in this chapter) are an impressive sign of this.

Political lock-ins describe situations where economic development proceeds along historical trajectories that are supported by cooperative relations between political actors, trade unions, employers associations and other embedding institutions. As a political lock-in, we identified the fact that in Stuttgart the economic and technology policy of the regional government was limited to modernization efforts within the core branches of the old economy. These branches received political support in order to stay competitive in highly competitive globalized markets. We have seen, however, that increasingly policy strategies have been introduced to develop economic and structural alternatives for the region in order to loosen regional dependence on established industrial clusters. In this case it must be said that the official policy on state and regional level has clearly changed focus and is attempting to support both the modernization of the old sectors as well as the promotion of new sectors by a policy of parallel institution building.

The development of strategies for institutional learning remains a challenge to be mastered. While many 'institutionally poor' economic regions seek to adopt the seemingly exemplary institutions of other countries, Stuttgart is faced with the challenge of restructuring and transforming an exceedingly rich institutional landscape. These

challenges put the well-considered (and concerted) modernization of regional institutions at the top of the agenda. The expedient further development of communication and cooperation promoting institutions is therefore a central precondition for the design and rejuvenation of innovation-friendly environments.

References

Ackermann, R. (2001) *Pfadabhängigkeit, Institutionen und Regelreform*. Tübingen: Mohr Siebeck.

Amin, A. and N. Thrift (1994) 'Living in the global', in A. Amin and N. Thrift (eds) *Globalization, Institutions, and Regional Development in Europe*. London and Oxford: University Press, pp. 1–22.

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

Braczyk, H.-J., G. Schienstock and B. Steffensen (1995) 'The region of Baden-Württemberg: A post-Fordist success story?', in E.J. Dittrich, G. Schmidt and R. Whitley (eds) *Industrial Transformation in Europe. Process and Contexts*. London and others: Sage, pp. 203–233.

Cooke, P. (1998) 'Introduction: Origins of the concept', in H.-J. Braczyk, P. Cooke and M. Heidenreich (eds) *Regional Innovation Systems*. London: UCL Press, pp. 2–25.

Cooke, P. and K. Morgan (1990) *Learning through networking: Regional innovation and the lessons of Baden-Württemberg*. University of Wales, College of Cardiff. Regional Industrial Research Report, No. 5.

David, P. (1985) 'Clio and the economics of QWERTY', *American Economic Review*, vol. 75, pp. 332–337.

Dosi, G. (1982) 'Technological paradigms and technological trajectories: A suggested interpretation of the determinants and directions of technological change', *Research Policy*, vol. 11, pp. 147–162.

Drache, D. (1996) 'Governance and public policy in a global economy. A report on jobs and investment strategies in Canada', Toronto, Canada: York University (unpublished).

Faust, M., H. Griupp, H. Hummel, G. Klee, T. Laube, W. Münzenmaier, C. Saul, U. Schmoch and C. Waldkircher-Heyne (1995) *Der Wirtschafts- und Forschungsstandort Baden-Württemberg. Potentiale und Perspektiven*. Studien zur Strukturforschung 19, Ifo Institut für Wirtschaftsforschung, Munich.

Gabriel, J. (1990) 'Innovation-oriented policy in regions with growth dynamics: Three winners in the process of structural change – A comparison of Baden-Württemberg, Massachusetts and Emilia-Romagna', in H.J. Evers and J. Allesch (eds) *Innovation and Regional Development Strategies. Instruments and Policy Coordination*. Berlin and New York: Springer, pp. 24–51.

Grabher, G. (1993) 'The weakness of strong ties. The lock-in of regional development in the Ruhr area', in G. Grabher (ed.) *The Embedded Firm. On the Socioeconomics of Industrial Networks*. London and New York: Routledge, pp. 255–277.

Hassink, R. (1992) 'Regional innovation policy: Case studies from the Ruhr area, Baden-Württemberg and the North East of England', *Nederlandse Geografische Studies* 145, Utrecht.

Heidenreich, M. and G. Krauss (1998) 'The Baden-Württemberg production and innovation regime. Past successes and new challenges', in H.-J. Braczyk, P. Cooke and M. Heidenreich (eds) *Regional Innovation Systems*. London: UCL Press, pp. 214–244.

Hughes, T.P. (1987) 'The evolution of large technical systems', in W.E. Bijker (ed.) *The Social Construction of Technological Systems*. Cambridge, MA: MIT Press, pp. 51–82.

ifo Institut für Wirtschaftsforschung (1995) *Der Wirtschafts- und Forschungsstandort Baden-Württemberg*. München (ifo).

Kaiser, W. (2007) 'Regionales Cluster oder globaler Knoten? Automobiltechnik im Raum Stuttgart', in H. Berghoff and J. Sydow (eds) *Unternehmerische Netzwerke*. Stuttgart: Kohlhammer, pp. 175–195.

Koschatzky, K. and A. Jappe (2004) Technologische Stärken und Schwächen der Region Stuttgart. Telekommunikation, IT und Optik verlieren an Boden – Umwelt-, Mess- und Regeltechnik im Aufwärtstrend. Stuttgart: Industrie- und Handelskammer Region Stuttgart.

Krauss, G. (1999) 'Les problèmes d'adaptation d'une économie régionale forte: Changement et inerties en Baden-Wuerttemberg', *Revue d'Economie Régionale et Urbaine* II, pp. 353–376.

Maier, H.E. (1989) 'Industrieentwicklung und Industriepolitik in Baden-Württemberg. Überlegungen zu den institutionellen Voraussetzungen differenzierter Qualitätsproduktion', in J. Hucke and H. Wollmann (eds) *Dezentrale Technologiepolitik? Technikförderung durch Bundesländer und Kommunen*. Basel, Boston and Berlin: Birkfellner, pp. 261–303.

Morgan, K. (1994) 'Reversing attrition? The auto cluster in Baden-Wuerttemberg', *Arbeitsbericht* Nr. 37, Stuttgart: Akademie für Technikfolgenabschätzung in Baden-Württemberg.

Pierson, P. and T. Skocpol (2002) 'Historical institutionalism in contemporary political science', in I. Katznelson and H.V. Miller (eds) *Political Science. State of the Discipline*. New York: W.W. Norton, pp. 693–721.

Rehfeld, D. (1995) 'Disintegration and reintegration of production clusters in the Ruhr area', in P. Cooke (ed.) *The Rise of the Rustbelt*. New York: St. Martin's, pp. 85–102.

Sabel, C.F. (1989) 'Flexible specialisation and the re-emergence of regional economies', in P. Hirst and J. Zeitlin (eds) *Reversing Industrial Decline? Industrial Structure and Policy in Britain and Her Competitors*. Oxford: Oxford University Press, pp. 17–70.

Saebetzki, A. (1994) 'Perspektiven der Beschäftigung im Maschinenbau, in der Elektrotechnik und im Straßenfahrzeugbau bis zum Jahr 2010', *ZEW-Wirtschaftsanalysen* I, vol. 2, pp. 78–95.

Schmitz, H. (1992) 'Industrial districts: Model and reality in Baden-Württemberg', in F. Pyke and W. Sengenberger (eds) *Industrial Districts and Local Economic Regeneration*. Geneva: ILO, pp. 87–121.

Semlinger, K. (1993) 'Economic development and industrial policy in Baden-Württemberg: Small firms in a benevolent environment', *European Planning Studies,* vol. I, pp. 435–463.

Semlinger, K. (1994) 'Industrial-district-Politik in Baden-Württemberg. Zwischen Neubesinnung und Neuanfang', *Arbeitsbericht* Nr. 39. Stuttgart: Akademie für Technikfolgenabschätzung in Baden-Württemberg.

Statistisches Landesamt Baden-Württemberg (1997) 'Region Stuttgart: Seit 1992 gingen 110000 Arbeitsplätze verloren', *Eildienst Statistisches Landesamt Baden-Württemberg*, Nr. 265/1997. Stuttgart: Statistisches Landesamt.

Storper, M. (1995) 'Regional technology coalitions: an essential dimension of national technology policy', *Research Policy,* vol. 24, pp. 895–911.

Thelen, K. (2004) 'How institutions evolve. Insights from comparative historical analysis', in J. Mahoney and D. Ruschemeyer (eds) *Comparative Historical Analysis in the Social Science.* Cambridge: Cambridge University Press, pp. 208–240.

Wallace, W. (1994) 'Rescue or retreat? The nation state in Europe, 1945–1993', *Political Studies,* vol. 42, pp. 52–76.

Part V

Path Dependence of Policy and Institutions

12

Explaining Path Dependence through Discourse Analysis: The Case of Seasonal Farm Workers in Germany*

Sebastian Hess, Daniela Kleinschmit, Ludwig Theuvsen, Stephan von Cramon-Taubadel and Ulrike Zschache

12.1 Introduction

This chapter introduces discourse analysis as a theoretical concept and an empirical methodology that may enable the endogenization of path-creating and path-breaking changes in conventional models of political path dependency. Economic criteria, such as rents created by a policy, do not always provide a comprehensive explanation of path-dependent political decisions. Discourse theory suggests that specific interpretative schemata and narratives, such as storylines in the mass media, heavily influence political discourse. Discourses themselves can actually exercise a constitutive power that constrains decision-making processes and, thus, influence the ensuing policy-creation path. Hence, discourses must be taken into account when political path dependency, path-breaking and path creation are analyzed. In this chapter we trace over time individual storylines that represent important elements of the discourse underlying the restriction of seasonal farm workers from central and eastern European countries in Germany. We illustrate how dominant speakers and their storylines have been interacting to shape this policy.

In the literature on the economics of technology, path dependencies are ascribed to positive feedback mechanisms that result in lock-in situations characterized by prohibitively high change costs that impede the dissemination of new technologies. These costs stem from high set-up costs, learning processes of technology users, coordination effects

(especially positive network externalities) and self-reinforcing expectations of actors (Arthur, 1994; Ackermann, 2001). More recently, political and organizational scientists have identified additional sources of path dependencies including cognitive and social causes of path-dependent processes (Pierson, 2000a; Schreyögg, Sydow and Koch, 2003). The concept of path-breaking describes a de-locking process initiated through exogenous shocks, external interventions (Ensminger, 1997; Gowan and Gunby, 1996) or cognitive or emotional processes (Schreyögg, Sydow and Koch, 2003). The term path creation was coined by Garud and Karnøe (2001a: 6): '[...] entrepreneurs may intentionally deviate from existing artefacts and relevance structures, fully aware they may be creating inefficiencies in the present, but also aware that such steps are required to create new futures. Such a process of mindful deviation lies at the heart of path creation.'

Path dependency within the political realm is marked by self-reinforcing feedback effects that alter the costs of switching from one policy regime to another (for instance, Kay, 2005, p. 558, 563). Pierson (2000a) identifies four elements of political systems that contribute to path dependency: the dominance of collective action in politics, the institutional density of politics, power asymmetries, and the complexity and opaqueness of political processes. These characteristics bring about that politicians tend to, or have to stick to political decisions made earlier. As a result of reaffirming dynamic processes, politics and institutions may get locked into situations that, once in place, become difficult to change (e.g., North, 1990). The Common Agricultural Policy (CAP) of the EU has frequently been cited and analyzed as an almost 'classical' example in this regard (Ackrill and Kay, 2006).

Pierson (2000a) notes that political path dependencies are far more complex and, due to a lack of easily measurable indicators, more difficult to analyze than cases of purely economic path dependency. Therefore, in the literature related to political science, Pierson (2000a, 2004) presents analyses that identify path dependence and explains why these dependencies exist within politics, yet without convincing and theoretically deeply rooted explanations of the reasons why certain – potentially inefficient – policies were introduced in the first place. Therefore, in economics, as well as in political science, the process of path dependence to date largely constitutes a research field with a theoretical framework that is just emerging and is still incomplete (Garud and Karnøe, 2001; Schreyögg, Sydow and Koch, 2003; Beyer 2006; Koch 2008). In addition, so far no adequate concept has been used that would allow the identification of the causes and circumstances under which

specific policies are introduced or altered, implying that the processes of political path dependency, path-breaking and path creation are not yet well understood.

This chapter proposes discourse analysis as a theoretical framework[1] and an empirical methodology that should be integrated into the analysis of political paths in order to reconstruct self-reinforcing feedback effects in politics in general and political institution-building and maintenance in particular. Further application of this theory with regard to path-dependent processes may enable predictions of political path dependencies, but may also show avenues for breaking or deviating from dominant paths. Taking the topic of seasonal farm workers from central and eastern European countries in Germany as an example, the processes of path dependence are empirically explored through a media analysis on this topic.

In the following section we summarize what established theoretical thought and empirical evidence can say about this policy issue and why the concept of path dependency in this context promises to fill certain gaps in conventional analysis. Then we introduce the concept of discourse analysis in connection with path dependency, before outlining a methodological framework that describes how the explanatory power of this concept can be empirically substantiated. Finally, we present and discuss our empirical results and draw some preliminary conclusions.

12.2 What explains the existence and persistence of inefficient agricultural policies? The case of seasonal farm workers

In (agricultural) economics, rent-seeking behavior (Krueger, 1974) and the associated activities of lobby groups often provide convincing explanations for the existence of protectionist policies that frequently turn out to have distortive effects (Alston, Norton and Pardey, 1995). Lobby groups seek to redistribute income in their own favor and, accordingly, lobby actively within politics. Assuming utility-maximizing behavior, the cost of the lobbying effort will be equal to or less than the volume of the actual rent involved (Krueger, 1974). In this context, agricultural policies have been analyzed by economists as well as political scientists for a long time and may be considered a classical example of redistributive policies that benefit the various farm lobby groups involved (e.g., Tangermann, 1976; Alston and James, 2002).

From the rational-choice perspective, politicians can be viewed as aiming to provide the best policies given various political constraints

(for instance, pressure arising from the activities of lobby groups; see Dixit and Romer, 2006). Alternatively, politicians and political institutions themselves can be seen as rent seekers (Olson, 1965: 'stationary and roving bandits') with selfish preferences, who are trying to maximize their own benefits rather than being motivated by the desire to provide in the best way possible for the public good. The analysis of distortive market policies common in agriculture describes the incidence of policy (Alston and James, 2002) typically as a failure to provide socially optimal outcomes due to a certain amount of redistribution of income in favor of certain lobby groups (Alston, Norton and Pardey, 1995; similar problems in reaching Pareto-optimal outcomes due to lock-in effects in agricultural cooperatives have been described by Svendsen, 2007).

These redistributive policies typically create economic rents. An economic rent can be defined as 'the payment to a factor of production over and above the minimum necessary to induce it to do its work' (Curry, Murphy and Schmitz, 1971, p. 758). Once an economic rent has been created and assigned to a group of beneficiaries, it can be argued that policymakers may already have induced political path dependency. Rents create a large potential for self-reinforcement due to the fact that beneficiaries will be unwilling to give up their privileges again (Krueger, 1974). However, a closer look at different definitions of path dependency, on the one hand, and individual agricultural policy measures, on the other, does not always clearly suggest that what is observed in reality necessarily implies anything more than a weak use of 'path dependency,' that is, not more than the general argument that 'history matters' (Ackrill and Kay, 2006). Therefore, a more specific look at individual policy fields provides better opportunities for analyzing processes of path creation and path dependence in the political sphere in more detail.

An example of a very specific, highly protective, and very persistent agricultural policy is the regulation of seasonal farm workers from central and eastern European countries (CEEC) who work each year in German agriculture (e.g., Gerdes, 2000). Although it can be traced back to the late nineteenth century, this policy seems to benefit neither farmers nor workers and is, at the same time, a perennial source of tension between lobbyists and politicians. In *de facto*, if not *de jure*, violation of the EU's common market, Germany and Austria continue to restrict the employment of workers from new EU member states in agriculture and neighboring economic sectors. Under the current regulation for seasonal farm workers from CEEC in Germany, farmers have to apply formally for a certain number of workers several months ahead

of the harvest season. Farmers have to prove that they really need these workers on their farms and that they are unable to fill vacant positions with German unemployed persons.

In general, farmers are currently granted only 80 per cent of the workers they have requested. Hence, in theory they are obliged to hire at least 20 per cent of their seasonal workforce on the German labor market. In practice, however, German workers are unable or unwilling to do the work in question. Therefore, a 20-per cent input restriction is imposed on labor-intensive agricultural products in Germany, or, in other words, an input quota equal to 80 per cent of total seasonal farm labor demand is in place. German farmers are all equally restricted by this 20 per cent cut of their labor demand. Compared to a scenario of free movement of workers, this policy therefore constitutes a politically induced market distortion with associated potential welfare losses (Hess, 2004). If rent seeking were the key motivation for the existence of this policy, at least one of the interest groups involved should clearly benefit in monetary terms. The following analysis shows that this is, in fact, not the case.

Input quotas typically limit the competitive market output of a farm product (Alston, Norten and Pardey, 1995). They also reduce the factor price equalization that would otherwise take place as high wages for farm labor in Germany attract low-priced workers from CEEC. This will, *ceteris paribus*, increase the price of labor as well as of the corresponding output product(s). Seasonal farm workers in Germany (both Germans and those in-quota workers from CEEC) clearly benefit through higher wages, while consumers of labor-intensive agricultural products clearly lose as a result of higher prices. The impact on farm enterprises that produce seasonal fruit and vegetable products is ambiguous as both output and input prices increase.

The political influence of seasonal workers from CEEC in Germany can be assumed to be low. Furthermore, workers in CEEC who do not get in-quota positions in Germany lose as a result of the policy. Hence, it is unlikely that this interest group has had an impact on the introduction and persistence of this policy, while almost no German seasonal farm workers exist (Gerdes, 2000). Consumers typically have little voice in agricultural market policy (price and trade policy measures) in the EU or specifically in Germany. It turns out that farmers' organizations are the strongest political opponents of seasonal farm worker regulations in Germany and lobby very actively against this policy. This indicates that of the two effects outlined above (increasing output and increasing input prices), the latter dominates and that farmers would be better off without the quota system.

German farm workers represented by the German labor union *Industriegewerkschaft Bauen-Agrar-Umwelt* (IG BAU) may fear incoming competitors driving down wages. Therefore, the union might have a strong incentive to lobby against seasonal farm workers from CEEC. However, since Germans are typically not willing to take seasonal jobs, there is no direct competition and, hence, German wages for year-round employees in agriculture will not be affected by the wages paid for seasonal farm hands. Thus, no direct rent-seeking effort by German labor unions is likely to be the driving force behind the politically induced reduction of farm-labor migration. On the contrary, from the union's perspective the CEEC workers can be considered as safeguards against societal pressure on union members to accept low-paid, arduous seasonal jobs in agriculture.

In theory, the quota on migrant farm labor from CEEC creates jobs for unemployed Germans representing 20 per cent of total seasonal farm labor demand. It would be reasonable to expect this group to have a vital interest in even more restrictive labor market protection and to be the real beneficiary of the rent generated by this policy. Instead, experience shows that the German labor administration initially had difficulties finding Germans who were willing and able to take on this work. Only after special training programs and additional monetary rewards were issued by the labor administration a few positions were filled by Germans. German farmers have frequently blamed policymakers for the resulting labor shortage. The lack of motivation for unemployed Germans to apply for jobs in agriculture indicates that rent seeking by this group is not a convincing explanation for the persistence of an inefficient agricultural policy (Hess, 2004).

Landowners are also frequently identified as the ultimate beneficiaries of protective agricultural policies. Although this is likely an important interest group with regard to the market protection of crops that are especially land intensive, in Germany less than 5 per cent of total farm land is cultivated with seasonal, labor-intensive crops (although these crops account for about 40 per cent of total sales from crops in Germany). Therefore, there are much more attractive policy arenas where landowners can invest in lobbying activities, for instance the emerging land-intensive production of bioenergies. Taking into account all the arguments discussed above, it is obviously hard to identify any specific interest group that clearly benefits in monetary terms from the existing policy that reduces farm labor employment. Nevertheless, the policy persists – a fact that obviously requires an alternative explanation.

12.3 Explaining path dependence of political processes through discourse analysis

The example of the restriction of seasonal farm workers in Germany shows that economic criteria such as rents created by a policy do not always provide a comprehensive explanation of path-dependent political decisions. Instead, there have to be different sources of positive, potentially self-reinforcing feedback processes in the political system. Therefore, for theoretically better substantiated and empirically sound explanations of the reasons for path-dependent processes, the framework needs to be enriched by an alternative theoretical concept. Discourse analysis is a rich theoretical concept, which can be applied to empirical analyses and offers the opportunity to understand social and political behavior *ex post*. The results of these analyses may in future also help to identify political path dependence *ex ante*.

Different concepts of discourse are used in policy analysis (Keller and Viehöver, 2006; Kerchner, 2006). This chapter focuses on the Foucaultian perspective of discourse practices. Foucault expects discourses to actively construct society along various dimensions and hypothesizes interdependencies between the discursive practices of a society and its institutions. Such practices, understood as texts, always draw upon and transform other contemporary and historically prior texts. Any given type of discursive practice is thus generated out of combinations of other analyses of collective knowledge orders and discursive practices. Therefore, a discourse is a bounded 'positive' field of statement accumulation implying at the same time that other possible statements, questions, perspectives, difficulties, and so on are excluded. These exclusions can be consolidated by institutions (Link and Link-Herr, 1990). In this meaning, discourses have a formative or constitutive power in structuring basic definitions and meanings that are later taken for granted.

Linking the concept of discourse analysis with the framework of path dependence leads to the assumption that discourses and their constitutive character can be seen as explanations for self-reinforcing processes in politics because political discourses are always built on historically prior texts; thus, the past strongly influences future political actions. The fact that definitions and meanings structured by discourses tend to be taken for granted creates psychological and institutional switching costs for policymakers and administrations. In this sense, the self-reinforcing mechanisms shaping political decision-making processes are similar to the ones introduced by Miller (1993) as an explanation for firms' strategic inertia and disability to change once successful but now obsolete firm policies.

Discourses take place at different levels: media, politics, science, literature, or administration. Identifying and explaining positive feedback processes in politics requires, of course, first and foremost the analysis of political as well as media discourse. While the first is a sign of political behavior, the latter provides a master forum including virtually everyone. It is '*the* major site of political contest because all of the players in the policy process *assume* its pervasive influence' (Ferree et al., 2002, p. 10).

With regard to path-dependent processes in politics, two main categories can be considered indicators of positive feedback processes as well as of path breaking or path creation: actors and storylines. Those who speak in the discourse usually represent the interests of collective actors. This position is very powerful, especially in the media. These speakers have the chance to give their interpretative pattern of a problem and, thus, actively shape the discourse. They can be connected to the storylines used. Thus, considering the diachronic dimension of the discourse, the prevalence of certain speakers and storylines can be interpreted as path dependence. In accordance with Pierson (2000a), we assume that the reasons for this prevalence (as well as for the path dependency of politics it reflects) also include, but are by far not restricted to, power asymmetries in the political system. Self-reinforcing processes exist due to the fact that the discourse gives meaning to events, enables thoughts and legitimizes actions. By doing so, discourse automatically excludes other potentials to speak, think and act (Landwehr, 2006, p. 109). This leads to a shortage of *acceptable* statements about reality. On the other hand, considerable changes in the composition of the speakers' ensemble or the emergence of new ideas underlying new storylines are indications of path-breaking and path-creation processes.

Thus, the case of the restriction of seasonal farm workers and the framework of path dependence and discourse theory leads to two main questions, which can be answered through an empirical investigation and subsequent discussion:

- Who are the actors influencing the public media discourse of seasonal farm workers with their interpretative pattern?
- Which storylines are used by actors shaping the media discourse?

12.4 Methodology

The empirical analysis is limited to the German *Süddeutsche Zeitung*. Quality newspapers like this one direct their messages to an elite

Table 12.1 Set of search words

Set of labor-related search words			**Set of farm-related search words**		
Arbeit (work)	Arbeiter (workers, employees)	Helfer (helpers)	Saison (season)	Ernte (harvest)	Landwirtschaft (agriculture, farming)
Wanderarbeiter (migrant worker)	Schwarzarbeiter (illicit worker)	Beschäftigung (employment)	Erdbeeren (strawberry)	Wein (vineyards, wine)	Gemüse (vegetables)
Arbeitslos (unemployed)	Sozialhilfe (social welfare provision)	Langzeit (long term)	Agrar (agri)	Spargel (asparagus)	Obst (fruit)
Jobs	Teilzeit (part-time)	Vorübergehend (temporary)			
Erwerbsperson (employed person)	Hartz IV (unemployed)	Arbeitskräfte (workers, employees)			
Beschäftigte (employees)	Geringfügig (marginal, part time)				

readership, especially to decision makers. Moreover, they are often quoted by other media. Therefore, the impact of this kind of newspaper is enormous (Gerhards, 1991, p. 44). The selection of articles used in the study was based on an online search using the keywords displayed in Table 12.1. An element of the first set of labor-related search words has been combined with each set of farm-related search words, and all elements from the resulting search word matrix have been applied to the Lexis-Nexis newspaper archive. The analysis covers the time period from 1991 to 2007. In this period, 198 articles on seasonal farm workers were published. These articles constitute the population of the exhaustive sample of the results presented in the following section.

All articles have been analyzed for their formal characteristics through a quantitative-qualitative content analysis. The basic category system, derived from media and discourse theories, is subdivided into groups of categories. For the media analysis presented here, the variables 'speaker' and 'storyline' are relevant. Possible attributes of the variable 'speaker' are subdivided to allow examination of which type of actor 'has a say' (speaks) in the national media public. Aspects of this variable are, for example, 'politics,' 'administration,' 'farmer association' and 'media.' These variables result from an explorative pretest of the empirical data.

The unit of analysis for categorizing the speaker is the statement. Statements, which are identified as relevant for coding, are single

verbal messages from actors speaking directly or indirectly in the article (Gerhards et al. 1998, p. 47). Thus, in one article, more than one message (statement) can be coded. Statements referring to seasonal farm workers are relevant for coding. Moreover, it has to be taken into account that the media themselves can appear as speaker. During the coding process, data are directly entered into a database through predefined forms. Also, during the analysis, continued cross-validation tests ensure the reliability of the persons conducting the coding.

12.5 Results

Figure 12.1 plots the number of utterances for important storylines over time and includes the number of unemployed Germans per year. Moreover, this figure indicates crucial political events regarding the regulation of seasonal labor from CEEC in Germany. The first important result is that the public debate (measured in terms of utterances per six months) intensifies remarkably whenever policymakers make decisions on work permits for seasonal workers from CEEC. The issue of seasonal farm labor gained significant media attention for the first time in 1998, when German politicians attempted to limit the total number of seasonal workers per farm to 85 per cent of a farm's seasonal

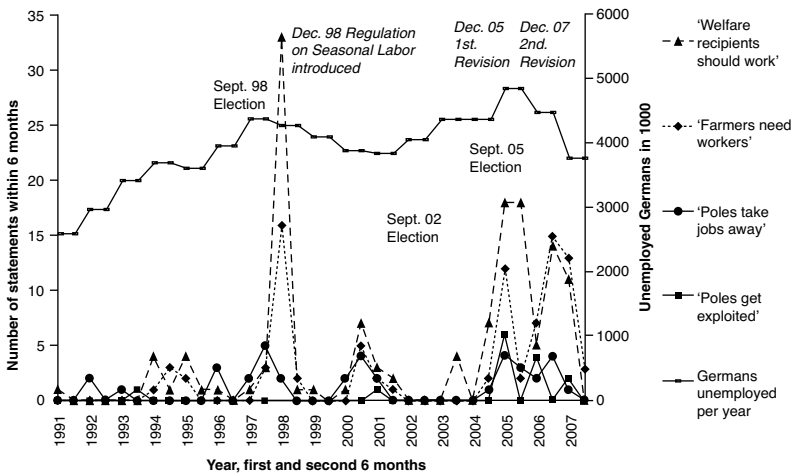


Figure 12.1 Important storylines over time in relation to the total number of unemployed Germans per year. Explanation: See text.

Source: Own analysis; unemployed statistic: Federal Department of Statistics, Germany.

CEEC employment in the year 1996. This regulation was first revised in 2005 under the new German coalition of conservative and social democratic parties. Their effort to re-restrict the number of foreign seasonal workers per farm to 80 per cent of the farm's total farm labor demand again resulted in a heated public discussion, as reflected in the related media coverage (cf. Figure 12.1). This can also be observed in the latest revision of the legal framework, which took place in December 2007. In this revision, the regulation was slightly relaxed: for instance, in regions with low domestic unemployment, farmers were now allowed to hire up to 90 per cent of their 1996 level of CEEC employment.

As for individual storylines, the storyline 'German welfare recipients should work in agriculture' occurs most often and peaks twice: in the federal government election years 1998 and 2005. In both years, domestic unemployment in Germany had experienced a strong increase, while it had been in decline prior to the election year 2002. Consequently, the topic of seasonal farm workers in German agriculture does not seem to have played a major role in the year 2002.

The second most important storyline is constituted by statements arguing that German farmers need seasonal workers. Obviously, this graph follows the first storyline closely, indicating that whenever arguments are raised to bring German unemployed people into seasonal jobs, German farmers and farmers' associations stress the fact that they cannot find enough qualified workers among Germans. Since the second half of 2006, this storyline has occurred even more often than the argument that Germans need to be brought into farm jobs. This can be explained by the fact that domestic unemployment declined sharply after 2005, which lessened pressure for politicians and the labor administration to 'find' jobs for Germans.

Table 12.2 displays the frequencies by which the different storylines depicted in Figure 12.1 have been used by different actor groups. It is evident that the most important storyline is that German unemployed persons should fill seasonal farm jobs first. Most actors address this argument, but reach slightly different conclusions. Policymakers and labor officials see the potential to reduce domestic unemployment, while farmers argue in this context that they have no objections to hiring Germans as long as they are as productive and reliable as Polish workers. Tables 12A.1 and 12A.2 (see Appendix) are supplementary and provide basically the same information as Table 12.2. However, Table 12A.1 presents the relative importance of storylines for each actor; in contrast, Table 12A.2 presents the relative importance of the related actors for each storyline.

Table 12.2 Number of times that a storyline has been addressed by an agent (count)

Storyline / Speaker	**German Policymakers**	**German Administration**	**Media**	**Farmers Associations**	**German Farmer**	**Single Worker/ German Unemployed**	**German Labor Union**	**Σ**
Foreign seasonal workers are exploited by German farmers.	0	0	12	3	0	1	0	16
German social welfare recipients should be required to work.	24	25	37	27	23	5	0	141
Polish workers introduce competition to the German labor market.	8	5	15	5	0	4	1	38
German labor-intensive farm products are selling well.	0	0	2	0	2	0	0	4
German farms need workers.	8	4	24	37	14	1	1	89
Other storylines	0	1	4	1	3	0	0	9
Σ	40	35	94	73	42	11	2	297

12.6 Discussion and conclusions

The empirical analysis of important storylines around the regulation of seasonal farm workers in Germany suggests that, in the media, policymakers and the German labor administration propose placing German unemployed people in seasonal jobs, while farmers stress these workers' limited productivity in comparison with Polish seasonal workers. In other words, in this instance, the motivation for labor-market regulation does not seem to be the desire to safeguard domestic wages, jobs, rents or other economic benefits. Instead, German officials use the media to confirm their interest in reducing domestic unemployment through placing Germans in seasonal jobs. This results from the fact that the domestic unemployment rate has been, and still is one of the most important issues in German politics, and voters continue to expect politicians to solve this problem. In turn, politicians are aware that, at election time, the public is judging their (anticipated) ability to reduce domestic unemployment. Therefore, as long as politicians seek to maximize votes, they must aim to appear in the media as competent reducers of domestic unemployment. The hypothesis that these attempts are especially fierce when domestic unemployment is on the rise and elections are near cannot be rejected. However, the extent to which other media support the same view and carry similar storylines remains to be examined.

Diagnosis of path dependence requires the identification of persistence, potential inefficiencies, environmental changes (which may contribute to the inefficiency of the locked-in path) *and* positive feedback mechanisms (which have created the lock-in situation and do not allow the implementation of a more efficient solution). These elements of path dependence can, in principle, also be identified in political processes (Pierson, 2000a). From a methodological point of view, this chapter has introduced discourse analysis as a theoretical concept and an empirical methodology that may enable the endogenization of political path creation, path dependencies and path breaking as a result of ongoing and changing discourses within societies.

This research has revealed that purely economic mechanisms, such as rent-seeking behavior, do not fully explain the emergence and persistence of the political regulations surrounding seasonal farm labor in Germany. Instead, the results suggest that public discourses exert an independent and important influence on the policymaking process by predefining the very nature of the political problem, the policy goals, and the instruments considered adequate. In this respect, public

discourses that restrict the definition of the reality and obscure possible alternative interpretations of the issue seem to be relevant sources of positive feedback processes.

Yet, on the basis of the data presented, it cannot be fully explained how self-reinforcing feedback processes develop in public discourses. The media coverage analyzed from the newspaper *Süddeutsche Zeitung* constitutes only a small part of the greater public discourse on seasonal farm work in Germany. In this regard, it can be assumed that general tendencies in the development of the discourse are difficult to explore. Therefore, in a further step, the current media analysis should be extended to more newspapers, thus capturing a wider spread of opinion and allowing public discourse on seasonal farm work to be described more comprehensively. In order to verify the hypothesis that discourses are powerful mechanisms that give rise to positive feedback processes in politics, further research should be able to demonstrate that individual discursive elements (such as storylines) become increasingly powerful over time until they eventually dominate the whole discussion on the issue. Furthermore, a growing convergence among statements by different actors and among different newspapers should be identifiable.

Referring to institutionalist theory, more recent discussions on path dependencies in the political realm underline the notion that key guiding assumptions, interpretative frameworks or 'shared mental models' (Denzau and North, 1994; see also Hall, 1993; Meyer et al., 1987) operating inside an 'organizational field' (DiMaggio and Powell, 1991) can be considered important factors that contribute to self-reinforcing processes in political decision-making (e.g., North, 1990, pp. 36–37 and 66–68; Pierson, 2000b; Mayntz, 2002, pp. 27–30; Thelen, 2003; Beyer, 2006). Path dependency in policymaking processes can be understood as a kind of 'context-bound rationality' shared among policy actors (Kay, 2005, p. 564; see also Mayntz, 2002, pp. 27–30). Due to the complexity of policy processes and the uncertain results of alternative policies, politicians or institutions act according to the predominant guiding assumptions. These collective interpretations are subject to self-reinforcing feedback processes because the formation of fundamental social understandings entails high start-up costs and learning effects. Moreover, socially shared interpretations and assumptions lead to adaptive expectations and coordination effects among social actors (Pierson, 2000b, p. 79; see also Thelen, 2003, p. 209 and pp. 216–217).

Disposing quotas for farm workers from CEEC would mean that politicians had to give up taken-for-granted assumptions of welfare policy, for example, that German unemployed are, at least in principle, willing to

take these jobs. The Agenda 2010 policy of the former German chancellor Gerhard Schröder illustrates how difficult it is to introduce a fundamentally new policy that includes a sharp break with formerly indisputable assumptions of social policy. 'Fordern und Fördern' (demand and support), one of Schröder's central catch phrases, for instance, reflected the new insight that at least sections of the unemployed may need higher incentives (as well as additional training) to seek jobs. This break with traditional basic assumptions of German (and especially social-democratic) welfare policy came at very high political costs and resulted in a stepwise pullback of the new government from this position. Furthermore, a broad public discussion about social justice came up and gave voice to the still dominant old paradigm. The political success of the newly founded party 'Die Linke' as well as a more pronounced social profile nearly all German parties have tried to develop over the last three years both demonstrate the huge power that the old discourse still exerts. All in all, the recent developments can be interpreted as a reaction of politicians to still dominant mental models in German politics and society with regard to the regulation of the labor market and an attempt to avoid political costs resulting from deviant standpoints.

In the perspective outlined above, German welfare policy seems to be subject to self-reinforcing processes due to dominant mental models and the inability of politicians to leave the framework defined by these models. In this sense, it can be hypothesized that the suboptimal regulation of seasonal farm labor in Germany became resistant to major change because politicians in the federal government responsible for welfare policy are influenced by the persisting dominant discourse on overall domestic unemployment.

Political decision-making processes are closely bound to discursive practices. In Western democracies, policymaking requires the presentation of convincing reasons to the public as well as to law-making bodies. Hence, discourse analysis appears to be a promising complement to conventional policy analysis that will help explain why some policies turn out to be path-dependent, how they become path-dependent and whether the discourses around certain policies can be approximated reasonably well through the qualitative and/or quantitative reconstruction of the speaking actors and corresponding storylines.

The results have interesting implications for politicians seeking to change policies as well as for political consultants. Both groups have to take into account the 'real world' effects of political decisions and the representation of old as well as new politics in public discourses. It may turn out that more or less path-dependent discourses are more

important determinants of politics than the real effects of these politics. Under such circumstances, political changes may be strongly dependent on the particular ability of actors to unlock discourses through processes of 'mindful deviation' (Garud and Karnøe, 2001a) or external intervention (Ensminger, 1997; Cowan and Gunby, 1996).

Future research should seek to establish a better understanding of the internal functioning of public discourses and strengthen the predictive power of discourse-analytic approaches in path research. So far, it has not been well understood under which conditions which actors participate with which intensity in a discourse. A better understanding of these processes would allow more successful influence on discourses over time and more successful influence on political decision-making processes.

Appendix

Table 12A.1 Relative importance of storylines for each agent (column share in per cent of column sum). For example, 'Out of all statements made by German policymakers, 60 per cent have addressed the fact that German welfare recipients should be required to work.'

Column Shares in %	*German Policy-makers*	*German Administration*	*Media*	*Farmers Associations*	*German Farmer*	*Single Worker/ German Unemp.*	*German Labor Union*
Foreign seasonal workers are exploited by German farmers.	0.0	0.0	12.8	4.1	0.0	9.1	0.0
German social welfare recipients should be required to work.	60.0	71.4	39.4	37.0	54.8	45.5	0.0
Polish workers introduce competition to German labor market.	20.0	14.3	16.0	6.8	0.0	36.4	50.0

Continued

Table 12A.1 Continued

Column Shares in %	***German Policy-makers***	***German Admin-istration***	***Media***	***Farmers Associa-tions***	***German Farmer***	***Single Worker/ German Unemp.***	***German Labor Union***
German labor-intensive farm prod-ucts are selling well.	0.0	0.0	2.1	0.0	4.8	0.0	0.0
German farms need workers.	20.0	11.4	25.5	50.7	33.3	9.1	50.0
Other storylines	0.0	2.9	4.3	1.4	7.1	0.0	0.0

Table 12A.2 Relative importance of agents for each storyline (row share in per cent of row sum). For example, 'Out of all statements for storyline 1 'exploitation', 75 per cent have been stated by the media.'

Row Shares in p.c.	***German Policy-makers***	***German Admin-istration***	***Media***	***Farmers Associa-tions***	***German Farmer***	***Single Worker/ German Unemp.***	***German Labor Union***
Foreign seasonal workers are exploited by German farmers.	0.0	0.0	75.0	18.8	0.0	6.3	0.0
German social welfare recipients should be required to work.	17.0	17.7	26.2	19.1	16.3	3.5	0.0
Polish workers introduce competition to German labor market.	21.1	13.2	39.5	13.2	0.0	10.5	2.6

Continued

Table 12A.2 Continued

Row Shares in p.c.	*German Policy-makers*	*German Admin-istration*	*Media*	*Farmers Associa-tions*	*German Farmer*	*Single Worker/ German Unemp.*	*German Labor Union*
German labor-intensive farm products are selling well.	0.0	0.0	50.0	0.0	50.0	0.0	0.0
German farms need workers.	9.0	4.5	27.0	41.6	15.7	1.1	1.1
Other storylines	0.0	11.1	44.4	11.1	33.3	0.0	0.0

Notes

* This study is part of a broader project on path-dependent processes in agricultural politics, supported by the German Research Foundation (DFG).

1. For a general understanding of discourse analysis as a theory see Wetherall et al., 2001. Our chapter focuses on discourse analysis in the Foucaultian sense (see Foucault, 2003; Hajer and Versteeg, 2005).

References

Ackermann, R. (2001) *Pfadabhängigkeit, Institutionen und Regelreform*. Tübingen: Mohr Siebeck.

Ackrill, R. and A. Kay (2006) 'Historical-institutionalist perspectives on the development of the EU budget system', *Journal of European Public Policy*, vol. 13 (1), pp. 113–133.

Alston, J. and J. James (2002) 'The incidence of agricultural policy', in B. Gardener and G. Rausser (eds) *Handbook of Agricultural Economics*, Vol. 2B Agricultural and Food Policy. Amsterdam: Elsevier, pp. 1689–1749.

Alston, J., G. Norton and P. Pardey (1995) *Science Under Scarcity*. Ithaca: Cornell University Press.

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

Beyer, J. (2006) *Pfadabhängigkeit: Über institutionelle Kontinuität, anfällige Stabilität und fundamentalen Wandel*. Frankfurt and New York: Campus Verlag.

Cowan, R. and P. Gunby (1996) 'Sprayed to death: Path dependence, lock-in and pest control strategies', *The Economic Journal*, vol. 106 (436), pp. 521–542.

Curry, J.M., J.A. Murphy and A. Schmitz (1971) 'The concept of economic surplus and its use in economic analysis', *The Economic Journal,* vol. 81 (324), pp. 741–799.

Denzau, A.D. and C.D. North (1994) 'Shared mental models: Ideologies and institutions', *Kyklos,* vol. 47 (1), pp. 3–31.

DiMaggio, P.J. and W.W. Powell (1991) 'The iron cage revisited: Institutional isomorphism and collective rationality in organizational fields', in W.W. Powell and P.J. DiMaggio (eds) *The New Institutionalism in Organizational Analysis*. Chicago: Univ. of Chicago Press, pp. 63–82.

Dixit, A. and T. Romer (2006) 'Political explanations of inefficient economic policies: An overview of some theoretical and empirical literature', paper presented at the IIPF Conference *Public Finance: Fifty Years of the Second Best and Beyond,* Paphos, Cyprus, August, 2006.

Ensminger, J. (1997) 'Changing property rights: Reconciling formal and informal rights to land in africa', in J.N. Drobak and J.V.C. Nye (eds) *The Frontiers of the New Institutional Economics*. San Diego: Academic Press, pp. 165–196.

Ferree, M.M., W. Gamson, J. Gerhards and D. Rucht (2002) *Shaping Abortion Discourse: Democracy and the Public Sphere in Germany and the United States*. Cambridge: Cambridge University Press.

Fischer, F. (2003) *Reframing Public Policy: Discursive Politics and Deliberative Practices*. Oxford: Oxford University Press.

Foucault, M. (2003) *Die Ordnung des Diskurses*. Frankfurt: Fischer.

Garud, R. and P. Karnøe (eds) (2001) *Path Dependence and Creation*. Mahwah, NJ and London: Lawrence Erlbaum Associates.

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

Gerdes, G. (2000) *Bedeutung der Arbeitskräftewanderung aus Mittel- und Osteuropa für den deutschen Agrarsektor.* Kiel: Christian-Albrechts-Universität, Agrar- und Ernährungswissenschaftliche Fakultät.

Gerhards, J. (1991) 'Die Macht der Massenmedien und die Demokratie: Empirische Befunde', WZB Paper FS III 91–108. Berlin: WZB.

Gerhards, J., F. Neidhardt and D. Rucht (1998) *Zwischen Palaver und Diskurs: Strukturen öffentlicher Meinungsbildung am Beispiel der deutschen Diskussion zur Abtreibung*. Opladen: Westdeutscher Verlag.

Hajer, M. and W. Versteeg (2003) 'A decade of discourse analysis of environmental politics: Achievement, challenges, perspectives', *Journal of Environmental Policy and Planning,* vol. 12 (3), pp. 175–184.

Hall, P.A. (1993) 'Policy paradigms, social learning, and the state', *Comparative Politics*, vol. 25 (3), pp. 275–296.

Hess, S. (2004) 'Die Beschäftigung mittel- und osteuropäischer Saisonarbeitskräfte in der deutschen Landwirtschaft', *Berichte über Landwirtschaft*, vol. 82 (4), pp. 602–624.

Kay, A. (2003) 'Path dependency and the CAP', *Journal of European Public Policy,* vol. 10 (3), pp. 405–420.

Kay, A. (2005) 'A critique of the use of path dependency in policy studies', *Public Administration,* vol. 83 (3), pp. 553–571.

Keller, R. and W. Viehöver (2006) 'Diskursanalyse', in J. Behnke, T. Gschwend and D. Schindler (eds) *Methoden der Politikwissenschaft: Neuere qualitative und quantitative Analyseverfahren*. Baden-Baden: Nomos, pp. 103–111.

Kerchner, B. (2006) 'Diskursanalyse in der Politikwissenschaft: Ein Forschungsüberblick', in B. Kerchner, and S. Schneider (eds) *Foucault: Diskursanalyse der Politik: Eine Einführung*. Wiesbaden: VS-Verlag, pp. 33–67.

Koch, J. (2008) 'Strategic paths and media management – A path dependency analysis of the German newspaper branch of high quality journalism', *Schmalenbach's Business Review*, vol. 60 (1), pp. 50–73.

Krueger, A. (1974) 'The political economy of the rent-seeking society', *The American Economic Review*, vol. 64 (3), pp. 291–303.

Landwehr, A. (2006) 'Diskursgeschichte als geschichte des politischen', in B. Kerchner and S. Schneider (eds) *Foucault: Diskursanalyse der Politik – Eine Einführung*. Wiesbaden: VS-Verlag für Sozialwissenschaften, pp. 104–122.

Link, J. and U. Link-Herr (1990) 'Diskurs-, interdiskurs- und literaturanalyse', *LiLi*, vol. 77, pp. 88–99.

Mahoney, J. (2000) 'Path dependence in historical sociology', *Theory and Society*, vol. 29, pp. 507–548.

Mayntz, R. (2002) 'Zur Theoriefähigkeit makro-sozialer Analysen', in R. Mayntz (ed.) *Akteure – Mechanismen – Modelle: Zur Theoriefähigkeit makro-sozialer Analysen*. Frankfurt: Campus, pp. 7–43.

Meyer, J.W., J. Boli and G.M. Thomas (1987) 'Ontology and rationalization in the western cultural account', in G.M. Thomas, J.W. Meyer, F.O. Ramirez and J. Boli (eds) *Institutional Structure. Constituting State, Society, and the Individual*. Newbury Park, CA: Sage, pp. 9–32.

Miller, D. (1993) 'The architecture of simplicity', *Academy of Management Review*, vol. 18 (1), pp. 116–138.

North, C.D. (1990) *Institutions, Institutional Change and Economic Performance*. Cambridge: Cambridge University Press.

Olson, M. (1965) *The Logic of Collective Action*. Cambridge, MA: Harvard University Press.

Pierson, P. (2000a) 'Increasing returns, path dependence, and the study of politics', *American Political Science Review*, vol. 94 (2), pp. 251–267.

Pierson, P. (2000b) 'Not just what but when: Timing and sequencing in political processes', *Studies in American Political Development*, vol. 14 (1), pp. 72–92.

Pierson, P. (2004) *Politics in Time: History, Institutions, and Social Analysis*. Princeton, Oxford: Princeton University Press.

Schreyögg, G., J. Sydow and J. Koch (2003) 'Organisatorische Pfade: von der Pfadabhängigkeit zur Pfadkreation?', in G. Schreyögg and J. Sydow (eds) *Managementforschung 13*. Wiesbaden: Gabler, pp. 257–294.

Svendsen, S.V. (2007) 'Lock-in of farmers in agricultural cooperatives: Reviving the effect of exit by means of constitutional amendments', in K. Karantininis and J. Nilsson (eds) *Vertical Markets and Cooperative Hierarchies: The Role of Cooperatives in the Agri-Food Industry*. Dordrecht: Springer, pp. 115–136.

Tangermann, S. (1976) 'Entwicklung von produktion, faktoreinsatz und wertschöpfung in der deutschen landwirtschaft seit 1950/51', *Agrarwirtschaft*, vol. 25, pp. 154–164.

Thelen, K. (2003) 'How institutions evolve', in J. Mahoney and D. Rueschemeyer (eds) *Comparative Historical Analysis in the Social Sciences*. Cambridge: Cambridge University Press, pp. 208–240.

Wetherell, M., S. Taylor and J. Yates (2001) *Discourse Theory and Practice. A Reader*. London: Sage.

13
Institutional Path Dependence: A Resistance to Controversies*

Kim van Nieuwaal

13.1 Introduction

Besides its classical focus on stability, path-dependence logic also has the potential to explain change. Institutional path dependence defined in terms of a resistance to controversies incorporates both stability and change. Sztompka's (1991) structuration-like theory of social becoming is used to argue that four levels of social reality are relevant in detecting where inertia and potential dynamics lie. It is argued that the ability to trigger, steer, settle or suppress controversies is an organizational capability (cf. Amit and Schoemaker, 1993, p. 35). The case of the Dutch small-fields policy is used to illustrate the example of two failed, yet significant, attempts to trigger a controversy. Both attempts were backed up by the logic that abandoning the current policy is both in the long and short run economically advantageous compared to maintaining the status quo. In other words, this case requires a path-dependence rationale that goes beyond the classical explanations for policy inertia. Applying the notion of the controversy points the finger at where and why lock-ins secure stability. In this case, it is particularly the lock-ins at the cognitive level of social structure that have suppressed the potential ability to trigger controversies which are aimed at altering the status quo.

This chapter is structured as follows: First, it will be argued that path-dependence thinking has the potential to explain both stability *and* change. Then, the notion of the controversy will be introduced as a contribution to path-dependence theory. Subsequently, it will be reported on how the empirical research to test the proposed notion has been carried out. Then, the Dutch small-fields policy will be dealt with, including two attempts to trigger the small-fields policy. It will

then be argued that we are dealing with a dormant controversy, as the lock-ins we discuss appear to be of a predominantly cognitive character. Subsequently, it will be argued that these lock-ins are sustained by specific ethics and values. The chapter will be completed with a discussion and conclusions, including some suggestions for further research.

13.2 Path dependence and the potential for explaining change

The notion of path dependence has gained an increasing interest within organization studies over the last decade (Sydow et al., 2005, p. 691). David's (1985, 1986) famous example of the QWERTY keyboard marks the beginning of a growing recognition of the concept's relevance within the field of research on organizational change. David pointed at the fact that market mechanisms do not necessarily lead to the most efficient solutions. In addition, as illustrated by the keyboard example, not all processes are easily reversible. The reason why we use the current computer keyboard today is not for efficiency's sake, but lies in past deliberations and decisions. Moreover, despite reasonable arguments for a different, more efficient configuration, a change process for that purpose is not likely to be feasible, due to the costs of immediate switching. The path-dependence logic thus explains why the maintenance of the status quo can prevail over change. The question now arises as to how path dependence can explain change itself, or if the concept is particularly suited to only understanding immobility.

Recently, attention has been given to unlocking path dependencies. For instance, 'Unlocking Organizations' was the theme of the 21st EGOS conference at Freie Universität Berlin in 2005. In a sub-theme on 'Path Dependence and Creation Processes in the Emergence of Markets, Technologies and Institutions', Jörg Sydow, Georg Schreyögg and Jochen Koch (2005, p. 20) have pointed at the reflexivity of actors in producing and reproducing paths and the resulting potential for change: 'The awareness of path dependence may trigger activities to loosen it.' Indeed, path dependence is a particularly interesting perspective because its focus on stability leaves open the question for change. In other words, path dependence is intriguing because change remains thinkable. This chapter intends to contribute to this promising development in path-dependence thinking.

In this chapter, it will be argued that path dependence can be understood as a resistance to controversies. Consequently, paths can be unlocked by means of controversies. The notion of controversy will

be elaborated on in terms of structuration theory, acknowledging the determining forces of social structures, on the one hand, and the voluntary potential of agents, on the other. The case of the Dutch small-fields policy, which is a crucial part of Dutch national gas policy, will be used to see if the notion can stand the empirical test. The field of public utilities is believed to be a worthwhile context for that matter, as it is characterized by the solidity involved in physical infrastructure and legislative frameworks, on the one hand, and the dynamics of the factor markets requiring flexible responsiveness, on the other.

13.3 The notion of controversy

The tension between change and continuity that path dependence incorporates resembles the duality of social structure of structuration-like theories (e.g. Giddens, 1984). Sztompka (1991), for instance, is not particularly interested in why societies remain stable, but rather why they change. According to structuration theorists, human and organizational behavior is guided by social structures, which at the same time exist because of the constant confirmation in human interaction. Sztompka distinguishes between four levels of social structure. The dialectical relationship between agency and structure, which drives the process of social becoming, occurs on one or more of these dimensions of social reality. 'Social structure must be seen as complex and integrated, consisting of several levels (or dimensions), but bound together by specific interlinkages' (Sztomka, 1991, p. 124). The ideal level deals with cognition: how is reality being perceived and interpreted? The normative level is about legitimacy: what is regarded as appropriate? The interactional level focuses on the network aspect: who is connected to whom? The opportunity level is the political approach to social reality: who has the power over which resources? These four levels of social structure can be analytically distinguished from each other, but in practice they cannot be separated.

Any moment in time, and each situation, is characterized by a particular conjunction of features of the four levels of social structure. It is the background against which agency can take place. Sztompka labels this as potential ability. Various scenarios for behavior are possible, but only one will be effectuated in practice. This is the manifested action. It is what is actually happening. Manifested action either confirms existing social structures, or questions and transforms them (see also Giddens, 1984). It is in the latter case that change could occur. The social structure needs to be defied on at least one of the four levels to potentially

reveal structural changes in society. Non-conformist behavior gives rise to a controversy: the current social structure, or parts thereof, is then debated. 'As a result, the definition of the situation could have become ambiguous. The existing norms and values might have become subject to debate. Interaction could have taken place in such a way that it circumvents the current interactional network. Opportunities might have been opened up or closed down for certain actors. Or any combination of these four possibilities could have taken place. The moment that such a controversy is resolved, either change has occurred or the old situation has been re-established' (Sminia, 2003, p. 1625). Any form of agency subjecting current social structures to a dispute for the first time will be called a trigger event. The outcome of the eventual settlement of the controversy determines whether social structure has been subjected to change or not. Regardless of the outcome, the existing social structures in combination with the available agents will supply another situation of potential ability. The story continues as the process of social becoming unfolds.

The controversy, as conceptualized above, acknowledges both the persistence of structures through time and the possibility to break with them. It thus connects with the intention of this chapter to elaborate on the potential for uncompelled agency in path-dependence logic. If change in social structures is dependent on controversies, and more particularly on the settlement of a controversy, then path dependence could also be unlocked by means of these controversies. As a consequence, path dependence can be explained as a resistance to controversies. Let us apply this definition to the QWERTY example. Path dependence here largely lies at the opportunity level and the interactional level. To begin with the latter, there are just too many people using the keyboard configuration to be able to switch to an alternate one instantly. At the opportunity level, which is not surprisingly connected to the interactional level in this case, there is also a financial lock-in: The number of keyboards and production lines spread all over the world encompasses such an immeasurable body of capital that modification would require a rather enormous financial expense to change, so nobody would, or could, invest in a changeover. Theoretically, however, it is doable. All users must then be reached and trained (interactional level) and the funds for that must be made available (opportunity level). Lock-ins at the other two levels of social structure are also thinkable, but these are not likely to involve the necessary conditions for inertia in this respect, albeit the overall relevance of each and every level in the social interaction process. In this particular case, it is safe to state that

the core of the controversy exists in the combination of the lock-ins at the two levels, delivering a barrier to thinkable alternative configurations that could come to the surface. Moreover, as time passes by, the barrier will only become bigger by an ever-increasing number of people using an ever-increasing number of keyboards. The temporal factor can thus enhance the argument that if change was to be opportune, it should have been implemented earlier, despite the good arguments for change that remain.

In this chapter, the notion of the controversy will be applied to the Dutch small-fields policy. The relevant question is why this policy has remained intact for more than three decades, despite good arguments against it. These arguments, as we will see, largely involve the opportunity level. From a financial-economic perspective, there would be sufficient conditions for an immediate unlocking of the small-fields path. Apparently, the lock-in lies elsewhere. Let us now look at the case more closely after having reported on the research methods, and then apply the proposed notion of the controversy to explain the institutional inertia of this particular policy.

13.4 Research methods

A case study has been conducted in order to generalize empirical findings to the theory (Yin, 1994). According to Eisenhardt (1989, p. 534), a case study is 'a research strategy which focuses on understanding the dynamics present in single settings'. Smaling (2003, p. 5) speaks of 'theory-carried generalization' to indicate 'that the theory functions as a carrier or as a vehicle'. The specific conceptualization of a controversy lies at the heart of the theory used for this study and can be considered a 'generative mechanism' (Tsoukas, 1989; Pettigrew, 1997). A generative mechanism is the motor that drives the process. As a consequence, for this study, it has to be detected whether the controversy indeed plays the role in the process that it has theoretically been attributed. In order to do so, desk research and qualitative interviews have been used as the main sources to reconstruct the narratives in terms of the proposed theoretical vocabulary. Interviews were held with professionals of the Ministry of Economic Affairs, the Ministry of Housing, Spatial Planning and Environment, mining company NAM, the Central Planning Bureau, RIKZ marine research institute, the Dutch Labor Party, gas distributor EBN and consultants IMSA Amsterdam. In addition, a database with almost 1,000 events on gas depletion in the Wadden Sea, which is an activity carried out by the Shell/Esso joint

venture and legitimized by the small-fields policy, was used to detect possible controversies.

13.5 The small-fields policy

The discovery in Groningen in 1959 of one of the largest natural gas fields in the world marked the beginning of the Dutch gas era. From then until 2004, gas revenues earned the state an estimated 159 billion Euros (VPRO/NPS, 2006). It has been one of the foundations on which the Dutch welfare state has been built. 'The Netherlands equals Belgium plus the gas revenues.' This is what prominent liberal party member Henk Vonhoff answered when asked what his estimates were of the economical significance of state revenues which had derived from the famous Groningen field (VPRO/NPS, 2006). Until the large-scale exploitation of that field from the early sixties onwards, the Dutch economy had been mostly relying on coal and oil as its main energy sources. Gas was being used prior to the discovery of the Groningen field on a low scale, but nobody would have thought at that time that gas would soon dominate national energy production and would even be exported. The effects on the economy and society as a whole would be even more impressive.

The first oil crisis forced the government to seriously reconsider its energy policy. The white chapter, which was presented by Minister of Finance Lubbers in 1974, stated that the depletion of the Groningen field had to be brought under control (MEZ, 1974, pp. 118–120). One of the means for long-term effective and efficient gas production, as suggested by the white chapter, was a 'small-fields policy'. The discovery and exploitation of additional fields had to be encouraged to benefit from the reservoir in Groningen as the cornerstone of Dutch energy production as long as possible. The Groningen field is unique for it is a relatively large and condensed reservoir. It is therefore a matter of opening the tap for a steady and firm flow of high-quality gas. Despite the low marginal costs involved and its size, the Groningen bubble is not exploited at full throttle, but only in conjunction with the smaller fields.

Gas from the smaller fields in the Netherlands is unsuitable for current devices in domestic households and industries. However, mixing it with the Groningen substance does not require any technical adjustments. The small-fields policy aims to benefit from the various reserves spread all over the country, in conjunction with the main stock. In addition, Groningen as a swing supplier contributes to disturbance prevention of

natural-gas delivery (Correljé, Van der Linde and Westerwoudt, 2003, p. 97). Particularly climatic conditions can cause a large variation in energy demand. Severe winters are well known for that. Its outstanding qualifications have turned the Groningen gas reservoir into an extremely valuable national resource. If it was not for the gas tax revenues, the famous delta works preventing the western part of the country from the sea, for example, would not have been possible.

13.5.1 First trigger attempt by the secretary-general of the Ministry of Economic Affairs

Despite its curriculum so far, the small-fields policy has not remained free of any scrutiny. However, critics of the policy have never really gotten through to the policy makers with their arguments. In the late 1990s, the secretary-general of the Ministry of Economic Affairs tried to make his ministry aware of the drawbacks of the small-fields policy, but without success. According to the secretary-general, the policy projects into the future production with low marginal cost, as is the case with the ready-to-flow Groningen gas, while the higher marginal costs involved in small fields are brought forward. Adhering to basic economic principles, he did not see the logic of exploiting the small-fields gas prior to the Groningen gas, which had lower marginal costs. 'All that needs to be done is to open the tap and the gas will flow abundantly. The small-fields policy does the exact opposite of what should be done: cheaper things first in the case of equal returns. Against that is the fact that earlier exhaustion implies loss of the flexibility benefits of the Groningen field 30 years from now if the exhaustion date is brought forward a few years; this loss occurs decades in the future, however, while the efficiency costs of the small-fields policy are incurred today.'[1]

The report he had in mind on evaluating that trade-off would never be finished though, at least not by the Ministry of Economic Affairs. The subordinates at the ministry could not relate to such an approach. In their view the small-field policy had a rather successful history and served a long-term goal of supply security. In the long run, as it is believed, exploring and exploiting the relatively expensive smaller fields in conjunction with the Groningen swing supplier will turn out to be cheaper than any other succession of short-term scenarios. In addition, the policy being part of a larger legal structure has attracted a lot of small-firm activities exploiting the little bubbles. Moreover, the state offers to participate in the case of favorable expectancies, joint ventures emerge, random depreciation is allowed and Gasunie, owning the infrastructural network, guarantees purchase. Nonetheless, not much later,

the civil servants had to defend the inertial small-fields policy to the scrutiny of the OESO by assuring that the gas revenues would not be used for structural expenses.

13.5.2 The second trigger attempt by the Central Planning Bureau

In 2006, the Central Planning Bureau published its report on the Dutch gas depletion policy. The conclusions were to a large extent in line with those of the secretary-general. The off-take guarantee for small-fields gas 'currently may outweigh the costs, but a further development of the gas market would reverse this picture' (Mulder and Zwart, 2006, p. 3). The report warns about the costs resulting from the reduced incentives for operators to respond optimally to short-term changes as is common in non-regulated markets. Regarding the Groningen cap, the report found that 'this measure is inefficient when the cap is binding, that is restricting the production from the Groningen field' (Mulder and Zwart, 2006, p. 3). In accordance with the claim of the secretary-general, the report emphasizes that the costs of capping Groningen follow from shifting returns to the future. The report was discussed in parliament but did not lead to any change of plans.

13.6 Dormant controversy

It has to be concluded that the small-fields policy has remarkably stood tall, despite the two occasional near-eruptions of the dormant controversy. The discussion apparently pops up every now and then in the ministerial echelons.[2] In June 2004, for instance, it led to a ceiling for the Groningen field of an average production maximum for the next decade. But more importantly, despite the arguments against it, it has remained a dormant controversy. Not even the environmentalists have taken up the challenge to try and oppose to it. The small-fields policy provides the legitimacy for gas depletion in the Dutch Wadden Sea, which is a highly controversial exercise. These wetlands are not only hiding several small fields, it is also a nature reserve of international importance, which is – according to many – threatened by gas depletion. Nonetheless, the environmentalists have not touched the small-fields policy at all. It has apparently become an institution trapped in a mode of path dependence. The longer it lasts, the harder it becomes to change things. Technical and legal aspects are only part of the explanation. It has to be concluded that the lock-ins sufficient to explain this institutional stability lie at the cognitive level. Taking into consideration what

has been said above, we now have reason to further investigate this particular level of social structure.

13.6.1 Lock-ins at the cognitive level

The small-fields policy is still going strong, but not because it is too expensive to 'unlock' it. On the contrary, we have just seen that, from a financial-economic point of view, rapid depletion is rather opportune, regardless of what is left in any of the remaining small fields. In other words, it is not the immediate costs of switching that prevent the path of unlocking – quite contrary to the QWERTY case. As the opportunity level does not supply sufficient explanations, the inertia seems to lie elsewhere. The attempts to trigger a controversy in order to subdue the small-fields policy in a dispute have never really gotten through. The reason for that is that the arguments used did not affect the cognitive level surrounding the policy. The Ministry of Economic Affairs has not rejected the argument as put forward by its former secretary-general and the Central Planning Bureau. The short-term financial benefits of ending the policy are just not a sufficient condition to change the minds of the policymakers. And that is what it is all about: 'the minds'. Something is stuck at the cognitive level.

At the Ministry of Economic Affairs, the small-fields policy appeals to a sense of responsibility that goes beyond enhancing economic trade. To them, the Groningen reservoir is a national strategic resource. This is not surprising, as this was exactly the reason why the small-fields policy was implemented in 1974. The Lubbers administration realized that the immense Groningen reservoir was something to treasure, now that the nuclear era did not seem to be the remedy for future energy problems. The small-fields policy itself was in fact a result of settlement of a controversy. For quite a while, the Netherlands had known that it was sitting on a giant gas bubble, but it did not immediately realize what that could mean. It did not only mean revenues, deriving from tax revenues and the international contracts that had been made, it also implied a certain independence regarding energy supply. This realization is, until this date, part of the culture at the Ministry of Economic Affairs, delivering a cognitive barrier to change, which is something worth elaborating on further in terms of path dependence, as will be done in the remainder of this chapter.

13.6.2 Other lock-ins

The lock-ins for the small-fields policy are centered around the cognitive level, but that does not mean that the relevance of the remaining

levels of social structure is absent. At the interactional level, there have always been close bonds between the Ministry of Economic Affairs and the mining company NAM, which operates most of the gas reserves in the country. NAM has recently started depletion of small-fields gas from under the Wadden Sea. It is expected to supply gas to the four largest Dutch cities for ten years (ECN, 2007). NAM and the Ministry of Economic Affairs have invested hundreds of millions of Euros in the Wadden gas and both are determined to take it out, despite the ongoing protests of environmentalists. We hereby touch upon the opportunity level as well. Although not comparable to the value of the Groningen gas, the small fields do harbor valuable resources. Due to the fact that the quality of gas varies per field and can only be processed through current household devices in conjunction with the Groningen gas, the termination of the small-fields policy would leave the small fields unused forever. In addition, at the normative level there are the expectations, investments and contracts based on thirty years of a particular policy. Again, obviously the levels cannot be seen separately.

13.6.3 Ethics and values

The dominance of the cognitive level in explaining the inertia of the small-fields policy draws attention to 'the network of ideas, beliefs, images, convictions about reality' (Sztompka, 1991, p. 124) involved in the small-fields policy. The sense of responsibility of the Ministry of Economic Affairs to treasure the national gas reserve as long and as well as possible resembles Weber's (1949) 'Verantwortungsethik' and directs us to the role of ethics and moral values. However, with this ethic of responsibility, or deontology, Weber prescribes an ethical alignment between agency and outcomes. It remains to be seen to what extent the causality at the cognitive level, portrayed by the small-fields advocates, indeed matches this criterion. In addition, responsibilities in terms of agency and effects are more a normative issue, rather than taking place at the cognitive level. It is perhaps more an example of Weber's 'Gesinnungsethik', which is the counterpart of his responsibility ethic. In line with Kantian ethics, this ethic of conviction connects intention with agency. Not surprisingly, convictions clearly belong to the cognitive level. Government encompasses certain 'convictions' that might not always stand the test of the empirical proof of the pudding. Nor are they always subdued to financial cost-benefit analyses. It is public values that are at stake. Jacobs (1992) has contrasted public values with private values. She speaks of two mutually excluding sets of values, respectively the 'guardian moral syndrome' versus the 'commercial

moral syndrome.' Van der Wal, De Graaf and Lasthuizen (2008), however, have indicated that some values are common to both the public and the private sphere. Nonetheless, some values remain unique to the public service. Parsimony and stability (e.g., De Bruijn and Dicke, 2006; Beck, Jørgensen and Bozeman, 2007) are such values that apply to the sentiments surrounding the small-fields policy.

The question now is not necessarily how these values and beliefs ended up at the cognitive level accompanying the small-fields policy – explanations for that can be found in the reasons for implementing the policy. But the real issue from a path-dependence perspective is that those values and beliefs got stuck there ever since, and, perhaps even more interestingly, that they managed to dominate the other three levels of social structure, thereby being a sufficient cause for institutional path dependence. The notion proposed here provides an explanation for the first issue by arguing that there has just not been a sufficient controversy to cause a shift at the cognitive level. In other words, the absence of a controversy serves as a self-reinforcing mechanism. The second issue can only partially be resolved by the proposed notion. Locked-in ethics and values can, apparently, prevail over social dynamics harboring potential or actual controversies. From that perspective, it could be concluded that cognition, hence perceptions, are perhaps harder to subject to a controversy and change than alliances, legitimacy and resources are.

Cognitive barriers are difficult to resolve because they are the result of complicated mechanisms of positive feedback loops, which are hard to unravel, let alone be interfered with. Values and beliefs are locked inside the intangible. They ended up in the minds of people through a process of internalization (cf. Wallis and Poulton, 2001). In our particular case, the implementation of the small-fields policy, including its underlying assumptions and argumentation, has been crucial in that respect. Routinization then sustains the cognitive frameworks as time goes by. The values and beliefs are transferred from one person to the other (interactional level). In addition, they are reflected in more tangible constructs such as infrastructural resources (opportunity level). And last but not least, as a consequence of the mentioned mechanisms, these values and beliefs become part of an organizational culture encompassing legitimacy frameworks (normative level). In our case study, we see how the values and beliefs that accompanied the origins of the small-fields policy have found their way into the minds of, amongst others, the policymakers, into plans, plants and pipelines, and into an organizational culture within the ministerial echelons. In other words, cognitive barriers are fed by iterative and cross-level

feedback mechanisms that not only involve the different dimensions of social structure, but become anchored within the intangible mindsets individually and (particularly) collectively. When cognition is involved, managing change will thus then find its biggest challenge at that specific level.

13.7 Discussion and conclusions

We have seen that the logic of path dependence provides an interesting perspective on the explanation for why the Dutch small-fields policy has remained intact for more than three decades. Institutional path dependence understood as a resistance to controversies, however, points the finger at specific lock-ins. In this particular case, it appeared that the cognitive level and its lock-ins dominated the other levels of social structure. The opportunity level, for instance, harbored the economic arguments, which appeal to classic path-dependence logic, but did not prevail. The proposed notion enriches path-dependence thinking with the realization that (a) stability can also be explained from the perspective of potential change, by arguing that the absence of controversies induces a self-reinforcing mechanism in itself, and (b) that several dimensions of social reality need to be distinguished in explaining both continuity and change.

The case of the Dutch small-fields policy has illustrated the example of failed attempts to trigger a controversy in order to change the status quo. For further research, it would be worth investigating cases where triggers have indeed been effectuated. Funnily enough, the proposed notion is based on the idea that classic path-dependence thinking explains continuity rather than change, yet the used example is one of institutional inertia. However, it is particularly the dialectical foundations of the notion that leave room for understanding stability as an absence of change. Another element that future research could incorporate is the active strategizing of agents to suppress potential controversies. In our study, we have seen that NAM, but also the Ministry of Economic Affairs, did not really have to put into practice strategies to prevent controversies from arising. This was not necessary, because lock-ins were already secured at the cognitive level. The proposed notion of controversy explicitly leaves room for interpreting the triggering, steering, settling and suppressing as an organizational capability which is believed to be a promising topic for future research.[3] Managing enterprises is managing paths. Future research could also pay attention to the relevance and dominance of one level of social structure compared

to the other. In our study, we have seen that the cognitive level overruled the other levels, but what are the conditions for that?

Notes

* Previous versions of this text have been presented at the EGPA Annual Conference in Madrid in September 2007 and at the Pfadkolleg Conference 'Studying Path Dependencies of Business, Institutions, and Technologies' in Berlin in February 2008. The author wishes to thank Mihaly Hogye, Ole Jess Olsen, Elke Schüßler, Jörg Sydow and Sweder van Wijnbergen for their specific and useful comments on earlier versions of the text.

1. Interview, 2 December 2005.
2. Personal correspondence.
3. www.institutionalsurvivalpath.com

References

Amit, R. and P.J.H. Schoemaker (1993) 'Strategic assets and organizational rent', *Strategic Management Journal*, vol. 14 (1), pp. 33–46.

Beck-Jørgensen, T. and B. Bozeman (2007) 'The public values universe: An inventory', *Administration & Society*, vol. 39 (3), pp. 354–381.

Correljé, A., C. van der Linde and T. Westerwoudt (2003) *Natural Gas in the Netherlands: From Cooperation to Competition*. Amsterdam: Oranje-Nassau Groep B.V.

David, P.A. (1985) 'Clio and the economics of QWERTY', *American Economic Review*, vol. 75 (2), pp. 332–337.

David, P.A. (1986) 'Understanding the economics of QWERTY: The necessity of history', in W.N. Parker (ed.) *Economic History and the Modern Economist*. Oxford: Blackwell, pp. 30–49.

de Bruijn, H. and W. Dicke (2006) 'Strategies for safeguarding public values in liberalized utility sectors', *Public Administration*, vol. 84 (3), pp. 717–735.

ECN (2007) *Energieverslag Nederland 2007*. Petten: ECN.

Eisenhardt, K.M. (1989) 'Building theories from case study research', *Academy of Management Review*, vol. 14, pp. 532–550.

Giddens, A. (1984) *The Constitution of Society. Outline of the Theory of Structuration*. Cambridge: Polity Press.

Jacobs, J. (1992) *Systems of Survival: A Dialogue on the Moral Foundations of Commerce and Politics*. New York: Random House.

MEZ (1974) '*Energiebeleid. Tweede Kamer, zitting, 1974–1975, 13 122 nrs. 1–6*'. The Hague: Ministerie van Economische Zaken.

Mulder, M. and G. Zwart (2006) '*Government Involvement in Liberalised Gas Markets: A Welfare-Economic Analysis of the Dutch Gas-depletion Policy*'. CPB document No. 110.

Pettigrew, A. (1997) 'What is processual analysis?' *Scandinavian Journal of Management*, vol. 13, pp. 337–348.

Smaling, A. (2003) 'Inductive, analogical, and communicative generalization', *International Journal of Qualitative Methods*, vol. 2, Article 5.

Sminia, H. (2003) 'The failure of the Sport7 TV-channel: Controversies in a business network', *Journal of Management Studies*, vol. 40 (7), pp. 1621–1649.

Sydow, J., G. Schreyögg and J. Koch (2005) 'Organizational paths: Path dependency and beyond', paper presented at the 21st EGOS Colloquium, Freie Universität Berlin, Germany, June 30–July 2, 2005 from http://www.pfadkolleg.de/

Tsoukas, H. (1989) 'Validity of ideographic research explanations', *Academy of Management Review*, vol. 14, pp. 551–561.

Sztompka, P. (1991) *Society in Action: The Theory of Social Becoming*. Cambridge: Polity Press.

van der Wal, Z., G. de Graaf and K. Lasthuizen (2008) 'What's valued most? A comparative empirical study on the differences and similarities between the organizational values of the public and private sector', *Public Administration*, vol. 86 (2), pp. 465–482.

VPRO/NPS (2006) *Andere Tijden*, Dutch TV show, 15 January 2006.

Wallis, K.C. and J.L. Poulton (2001) *Internalization: The Origins and Construction of Internal Reality*. Buckingham and Philadelphia: Open University Press.

Weber, M. and E. Shils (1949) *The Methodology of the Social Sciences*. Glencoe: Free Press.

Yin, R.K. (1994) *Case Study Research. Designs and Methods*. Thousand Oaks: Sage Publications, Inc.

Part VI

Perspectives on a Theory of Path Dependence

14

Strategic Consequences of Co-evolutionary Lock-in: Insights from a Longitudinal Process Study

Robert A. Burgelman

14.1 Introduction

Modern economic theory recognizes that competition between alternative technologies is under certain conditions characterized by 'increasing returns to adoption' (e.g., Arthur, 1987, 1989, 1994). Arthur points out that the process involved is '... *non-ergodic* – or more informally we can say that it is *path dependent* in the sense that the outcome depends on the way in which adoptions builds up, that is, on the path the process takes' (1987, p. 438). An important aspect of the competing-technologies adoption process is that it is '... inherently unstable, and it can be swayed by the cumulation of small historical events, or small heterogeneities, or small differences in timing' (1987, p. 438). And, an important theoretical implication of this is that 'What we have in this simple model is order (the eventual adoption-share outcome) emerging from fluctuation (the inherent randomness in the arrival sequence). In modern terminology, our competing-technologies adoption process is therefore a *self-organizing process*' (1987, p. 438). As such, models of increasing returns to adoption and path dependence can be related to theory about complexity and chaos in the physical sciences (e.g., Kauffman, 1993; Prigogine, 1981, 1996; Nicolis and Prigogine, 1989).

Also, and crucial for the research approach described in this paper, since Arthur and his colleagues as mathematical economists face a non-Laplacian world, their modelling strategy must assume that the different pattern of small events that can lead to very different outcomes is unknown at the outset, and they must therefore treat these as random (Arthur, 1987, p. 438). Related to this, the historian John Lewis Gaddis

(2002) refers to the palaeontologist Stephen Jay Gould (e.g., 2002), who has highlighted the critical role played by contingency in evolutionary processes, and points out that 'rerunning the tape,' if that were possible, would produce different results. By implication: 'Any reliance on reductionism to simplify the past in order to anticipate the future becomes unworkable in these situations, and we're back to the old-fashioned historical narrative (...) as a more sophisticated research tool than most social scientists – indeed than most historians – have yet realized' (Gaddis, 2002, p. 81). Not surprising, therefore, economic historian Paul David's (1985) narrative of how and why the QWERTY configuration of the typewriter keyboard has remained the global standard for more than a century remains the paragon for describing and conceptualizing path dependence: a non-rational choice (in the sense of it not being the optimal outcome of a rational, comprehensive decision-making process) at the start of the diffusion process became amplified through network externalities (one source of increasing returns to adoption) and resulted in 'lock-in,' which according to Arthur means that '...an ever larger boost to the payoff of the excluded technology would be needed to resuscitate it' (1987, p. 438).

Finally, and also crucial for the research approach adopted in this paper, while studies of increasing returns to adoption and path dependence have so far been focused primarily on the development and diffusion of technologies, the broader theoretical frameworks of complexity theory and chaos theory are increasingly applied in social science (e.g., Axelrod and Cohen, 2000; David, 1990; North, 1990) and history (e.g., Gaddis, 2002). Management scholars have also attempted to introduce some of these theoretical ideas into administrative science (e.g., Burgelman, 1983b; Burgelman and Grove, 2007; Thietart and Forgues, 1995; Brown and Eisenhardt, 1997; Levinthal, 1997; McKelvey, 1997; Anderson, 1999; Tsoukas and Chia, 2002; Meyer, Gaba, and Colwell, 2005). Discussing some of the original theorists' work Gould, however, cautions against '...any pure theoretician's claim that success in modeling logically entails reification in nature' (2002, p. 927).

In this paper, I adopt Gaddis's methodology of historical narrative and heed Gould's caution. Informed by evolutionary organization theory (e.g., Aldrich, 1989; Baum and McKelvey, 1999) and based on longitudinal field research of Andy Grove's highly successful strategic leadership as Intel's CEO during 1987–1998 (Burgelman, 2002a, 2002b), I identify co-evolutionary lock-in and examine its consequences as a potentially interesting manifestation of the effect of increasing returns to adoption and path dependence at the *firm level of analysis*. Since the

challenges posed by increasing returns to adoption and path dependence at the firm level of analysis unavoidably need to be addressed by a company's strategy-making process, I examine the consequences of co-evolutionary lock-in within the framework of induced and autonomous strategy processes (Burgelman, 1983a, 1991). In particular, I examine the effect of co-evolutionary lock-in on top management's capacity to balance induced and autonomous strategy processes so as to maximize both 'fitness;' that is, adaptation to the current environment, and 'evolvability;' that is, ability to adapt to a changing environment and/or to seek out new viable environments (Kauffman, 1993; Gould, 2002), which is at the heart of sustaining corporate longevity (Burgelman and Grove, 2007).

14.2 Strategy-making and organizational evolution

14.2.1 Internal ecology of strategy-making

Evolutionary organization theory has grown into a wide variety of applications of Campbell's (1960) original discussion of variation, selection and retention processes (e.g., Aldrich, 1999; Baum and McKelvey, 1999). Earlier evolutionary research concerning Intel's transformation from a memory company into a microprocessor company (Burgelman, 1991, 1994; Grove, 1996) suggested an internal ecology view of strategy making, which was conceptualized in terms of induced and autonomous strategy processes (Burgelman, 1983a, 1991). Figure 14.1 shows the evolutionary framework of the strategy-making process in established companies.

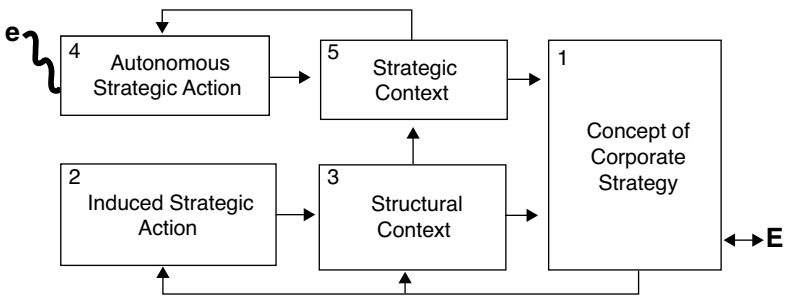


Figure 14.1 An evolutionary framework of the strategy-making process in established companies

Source: (Burgelman, 2002, p. 9).

As shown in Figure 14.1, induced strategic action exploits opportunities that are within the scope of a company's current strategy and extend it further in its current product-market environment (E). Autonomous strategic action, which emerges spontaneously from an organization's dynamic capabilities (Burgelman, 1983, 1991; Burgelman and Siegel, 2008), exploits opportunities that emerge through exploration outside of the scope of the current strategy and provides the basis for entering into new product-market environments (e). Intel's strategy making resembled an internal ecology in which induced (memory related) and autonomous (microprocessor related) initiatives competed for the company's scarce resources based on their success in the external competitive environment. Strategy-making viewed through the lens of internal ecological processes provided insight in forces driving organizational change that were not contemplated in the original theoretical articulations of organizational ecology (Hannan and Freeman, 1977, 1984).

The organizational ecology perspective suggests that organizational change must be understood at the level of entire populations of similar organizations, and as the result of replacement and selection rather than adaptation. Incumbent companies fail in the face of environmental change because inertia prevents them from adapting and are replaced with new ones that do different things or the same things differently ('better,' in the eyes of the majority of customers). The study of Intel's exit from the DRAM business adds empirical evidence in support of organizational ecology. Organizational ecology, however, leaves little room for adaptation based on strategy. Yet, strategy-making processes clearly helped Intel transform itself from a memory company into a microprocessor company, thereby preventing its demise. Hence, organizational ecology does not always provide a complete explanation of organizational change.

Established companies continue to remain subject to the selection force of the external environment. Many, even very large ones, do in fact succumb to it in the long run (Burgelman and Grove, 2007). But established companies have also gained the opportunity to substitute, to some extent, internal selection for external selection. This is the central idea of the internal ecology model of strategy making. An established company can be viewed as an ecological system in its own right, and its survival and continued success depend on the functioning of its internal ecology of strategy-making, which constitutes an *adaptive organizational capability*. While ecological processes at the level of organizational populations (industries) involve organizational founding and disbanding rates, the internal ecology of strategy-making involves entering new

businesses and exiting from failing businesses over time. Different parts of the internal ecology of strategy making can be linked to different forms of adaptation, and this helps reconcile opposing ideas about various consequences of strategic change (Burgelman and Grove, 2007).

14.2.2 Co-evolutionary lock-in

With strategy making as an adaptive organizational capability in mind, further longitudinal field research examined Andy Grove's performance as CEO within the extended historical context of Intel as an evolving system (Burgelman, 2002a, 2002b). Grove stepped into the CEO position at the time of a critical transition for Intel and he left the CEO position when a new critical transition was imminent. This research documented how Grove turned Intel's internal ecology of strategy making into more of a 'Rational Actor' (Allison and Zelikow, 1990; Burgelman, 2002a) approach to strategy-making. Among the first to understand the strategic implications of the 'horizontalization' of the PC industry (Grove, 1993; Farrell et al., 1998) and the associated 'increasing returns to adoption' (Arthur, 1987), Grove was able to 'vectorize' Intel's strategic action; that is, he was able to align the full magnitude of the strategic forces at Intel's disposal in the same direction and drive the evolution of the PC industry during the time he was CEO. The positive environmental feedback associated with Grove's successful strategy vector, however, caused *co-evolutionary lock-in*, a powerful but little noticed process in the literature on organizational adaptation. Co-evolutionary lock-in helped shed light on how a company's strategy making becomes specialized for its product-market environment (Barnett, 1997) and on the potential limitations of 'time-paced evolution' (Gersick, 1994; Brown and Eisenhardt, 1997). It also showed how co-evolutionary lock-in may affect the effectiveness of the autonomous strategy process, and thus a company's internal ecology of strategy making and its long-term adaptive capability.

In retrospect it is clear that the longitudinal field research of Intel as an evolving system (Burgelman, 2002a, 2002b) served the role of historical narrative, highlighted by Gaddis (2002), in documenting new forms of path dependence as the company was able to become the sole source supplier of microprocessors for the fast-growing PC industry. It took grounded theorizing (Glaser and Strauss, 1967), however, to identify and conceptualize the phenomenon of co-evolutionary lock-in and the new forms of strategic inertia that were its unintended consequence. These new forms of strategic inertia draw attention to the potential limitations of 'guided evolution' (Lovas and Ghoshal, 2000). They also

help examine further the importance of strategic context determination in maintaining a balance between induced and autonomous strategy processes, and how these processes may be linked to exploitation and exploration in organizational learning (March, 1991) and to theory about organizational ambidexterity (O'Reilly and Tushman, 2008).

14.3 Strategic consequences of co-evolutionary lock-in

14.3.1 Co-evolutionary lock-in begets dependence

As the sole source of the highest value component of PCs, which were becoming commodity products, Intel became the driving force in its co-evolution with the PC market segment. Intel's success created a positive feedback loop that made it increasingly able to appropriate the available rents in the PC industry. This asymmetry, however, also required Intel to make more and more of the investments necessary to secure timely industry adoption of its ever more powerful microprocessors, including manufacturing investments, R&D investments in enabling technologies to help OEM customers upgrade other components of the PC system, co-marketing investments with OEM customers, and investments in complementors (companies that provide complementary products that enhance the value of Intel's microprocessors for the end users). In some ways, these strategic investments served Intel to control its external environment (Pfeffer and Salancik, 1978). However, while these complementary strategic thrusts made Intel less dependent on the strategies of other players, they increased the dependence of its economic fortunes on the PC market segment. Also, as James March points out, strategic dominance may entangle a company in a system of relationships that reduce its freedom of action: 'You can have autonomy or you can have power but you cannot have both. Power depends on linkages and linkages destroy autonomy' (personal communication). Intel's OEM relationships, for instance, *de facto* reduced Intel's freedom to forward integrate further into systems products. And the relationship with Microsoft *de facto* reduced its freedom to move into software products.

> Insight 1. Co-evolutionary lock-in of a company's strategy with its product-market environment causes dependence and reduces its strategic freedom in some ways.

14.3.2 Time-paced strategy reinforces co-evolutionary lock-in

Some authors have suggested that time-paced strategy is a powerful alternative to event-paced strategy (Gersick, 1994; Brown and Eisenhardt,

1997). Time-paced strategy, in principle, allows a company to dictate the pace of strategic change that other players – customers, competitors, suppliers, and complementors – must adhere to. As a driving force of the PC market segment, Intel was able to influence the pace of change. Nevertheless, Intel's time-paced strategy did not simply reflect the company's ability to unilaterally impose its strategic intent on the product-market environment. Andy Grove had learned that there was a natural adoption cycle in the PC market segment, with a period of about three years between the maximum ramp-up for different microprocessor generations. He also understood that Intel could not expect to change this much. At the same time, having put in place the competencies and support infrastructure to deliver new generations of microprocessors to the PC market segment, there were both external expectations that Intel would deliver as well as an internal drive to do so.

> Insight 2. A successful time-paced strategy ties a company to the natural product adoption cycle of its product-market environment and reinforces co-evolutionary lock-in.

14.3.3 Competitive intensity reinforces co-evolutionary lock-in

Some authors have suggested that surviving competitors in intensely competitive industries may be quite strong but also highly specialized in their competitive repertoire (Barnett and Hansen, 1996; Barnett, 1997). Intel had been able to win the two defining battles – against other IA suppliers and against RISC – that Andy Grove had identified in late 1993 by developing competencies that were highly specialized for the PC market segment and gave the company the opportunity to pace the race through product leadership. Intel's introduction of the Celeron processor in 1998 to counter AMD at the low end of the PC market segment testified to the company's continued capacity to make quick adjustments that reinforced its competitive strategy. The need for a crash effort to introduce the Celeron processor, however, suggests that Intel, while innovating at a high rate, had begun to produce innovations – for instance, microprocessors based on design-to-performance rather than design-to-cost – that were less in tune with evolving environmental demands (Sorenson and Stuart, 2000). Intel seemed to have difficulties recognizing that the importance of the external selection environment relative to the internal selection environment was increasing toward the end of Grove's tenure as CEO (Sorenson, 2000). Intel's difficulties in this respect seem consistent with the observation that in successful organizations there will be a natural tendency for internal

selective-retentive processes to dominate external ones (Miller, 1999, p. 94). Co-evolutionary lock-in may thus be a potentially interesting extension and further elucidation of structural inertia (Hannan and Freeman, 1984).

> Insight 3. The stronger a company's competitive intensity the more its strategy becomes specialized to its product-market environment, which reinforces coevolutionary lock-in.

14.3.4 Co-evolutionary lock-in engenders misapplying the core business strategic logic

The success leading to co-evolutionary lock-in may be associated with a heightened level of confidence on the part of the CEO in the core business strategic logic. This carries potential danger if the CEO decides to apply the strategic logic of the core business to new business development efforts. Andy Grove's direct involvement in a PC-based videoconferencing venture (called ProShare), for instance, made it difficult for the middle-level executive in charge to develop a strategy that was appropriate for the new business and to act in accordance with an objective analysis of the situation. Grove's strategic intent for ProShare supports Audia et al.'s (2001) suggestion that success may increase a decision maker's feelings of self-efficacy and introduce an element of complacency that is best understood in terms of the strength of the decision maker's beliefs in the validity of the current strategy. It also supports Miller's (1994) finding that decision-making styles tend to be more 'extreme' during periods following success than during periods following poor or mediocre performance. Grove eventually came to realize this, but his strong involvement early on, before major market and technical uncertainties had been reduced, led to premature commitment of significant amounts of corporate resources. Once the CEO had committed to the new business it was difficult to avoid escalation of commitment and to scale down or extricate the company from the failing business.

> Insight 4. Co-evolutionary lock-in reinforces top management's confidence in the core business strategic logic as a driver of new business development, and impedes discovery of an appropriate strategy for the new business, precipitates premature commitment of significant amounts of corporate resources, and delays scale down or exit.

While insight 4 concerns companies that have achieved strategic dominance in their product-market environments with narrow

business strategies, it nevertheless raises important questions about the limitations of using top management's strategic intent as a means for 'guided evolution' (Lovas and Ghoshal, 2000). Within Intel's induced strategy process guided evolution worked fine: many new projects related to the core microprocessor business strategic logic were useful variations that were effectively selected and retained. However, when Andy Grove tried to shape what appeared to be new core business-related variations, but turned out to be quite different with the core business strategic logic, the result was misguided evolution. Insight 4 thus suggests that an internal ecology model encompassing induced and autonomous strategy processes that leaves room for middle and senior levels of management to help shape new organizational-level strategic intent may offer a richer and more complete view of strategic leadership than the rational actor model in which strategic intent emanates only from the CEO.

14.3.5 Co-evolutionary lock-in impedes non-core new business development

Although much of Intel's R&D investments went into technologies that complemented the microprocessor and thereby offered opportunities to launch other new businesses, the company rarely attempted to do so. An important reason for this was that any technology advance that enriched the PC environment was likely to create more demand for microprocessors, which commanded very high margins given Intel's sole source position. Thus, it was generally more valuable in the short run for Intel to give away technology and quickly disseminate it in the market, rather than try to build a business around it. This indicates another powerful effect of co-evolutionary lock-in. As Intel's extraordinary lucrative core business continued to grow very fast in the mid-1990s, Andy Grove increasingly considered non-core business opportunities as distractions. Also, Grove's effective communication with the business press and with stock market analysts reinforced Intel's identity as a highly focused microprocessor company (Zuckerman, 2000). Consequently, it became increasingly difficult for non-core new businesses to command top management attention and obtain corporate resources.

> Insight 5. Co-evolutionary lock-in requires a company to make industry-enabling investments that advance its core business, and creates difficulties for non-core new business development efforts to sustain consistent corporate support.

Insight 5 is not incompatible with insight 4. Insight 4 concerns 'induced' new business opportunities that are perceived by top management to be directly derived from the core business, such as the PC videoconferencing (ProShare) strategic initiative championed by Andy Gove; insight 5, on the other hand, concerns 'autonomous' new business opportunities that are perceived to be outside of the core business, such as Intel's networking business in the early 1990s (Burgelman, 2002a, 2002b).

14.3.6 Co-evolutionary lock-in hinders strategic context determination

Co-evolutionary lock-in increasingly focused Intel's structural context on execution of the core business strategy as well as on quick responses to perceived threats such as the Network Computer and the rapid growth of the low-end part of the PC market segment. It was less able, however, to deal with non-core new business development. Strategic planning was almost exclusively focused on the core business. New general management talent was not easily developed in Intel's matrix organization. Resource allocation favored the core business, and new businesses were constantly in danger of experiencing random shocks when critical resources were taken away to cope with a perceived threat to the core business. The measurement and reward system was unforgiving for deviations from objectives, even though new business strategies require such flexibility. While many new ideas continued to emerge, the structural context increasingly attuned to executing the core business strategy made activation of the process of strategic context determination for new businesses difficult.

> Insight 6. Co-evolutionary lock-in reinforces a company's structural context for executing the core business strategy and makes it harder to activate strategic context determination processes for non-core new business development.

14.4 Discussion and conclusion

Previous findings based on a study of Intel's transformation from a memory company into a microprocessor company (Burgelman, 1991; 1994) supported the insight that companies that are successful over long periods of time maintain top-driven strategic intent (through the induced strategy process) while *simultaneously* maintaining bottom-up driven strategic renewal (through the autonomous strategy process). Recent efforts by researchers of the modern economic theory of the

firm to formalize parts of the induced and autonomous strategy processes framework seem to support this insight (Rotemberg and Saloner, 2000). The study of Andy Grove's tenure as CEO, however, shows the difficulty of maintaining an appropriate balance between induced and autonomous strategy processes in the face of co-evolutionary lock-in (Burgelman, 2002b), itself the result of the extraordinary success of a narrow business strategy (Rotemberg and Saloner, 1994) pursued through the induced strategy process.

Simultaneously maintaining induced (variance reducing) and autonomous (variance increasing) strategy processes may involve similar difficulties as maintaining a balance between exploitation and exploration processes in organizational learning (March, 1991). Both processes compete for limited resources and companies necessarily make tradeoffs between them. Intel's induced strategy process during Andy Grove's tenure as CEO was maximally concerned with exploitation. While Grove successfully transformed Intel's internal ecology of strategy-making into the rational actor approach to strategy-making, toward the end of his tenure it had become clear that Intel's future as a growth company would also depend on autonomous new business development and that the strategies for new businesses might have to be defined by general managers closer to the frontline. Yet, the extreme emphasis on the induced strategy process during Grove's tenure had reduced the effectiveness of the company's autonomous strategy process.

The analysis of the consequences of co-evolutionary lock-in in this paper thus confirms that continued adaptation depends critically on strategic leadership able to maintain an appropriate balance between induced and autonomous strategy processes in a company's internal ecology of strategy making. This strategic leadership imperative may also help resolve the seeming contradiction between a study of corporations that attributes exceptional long-term success to leadership style (Collins, 2001), and the more common proposition that strategy is the determinant of long-term performance. While Collins's study found, surprisingly, that strategy does not play a decisive role, the paradox is resolved if what sets apart leaders of 'great' companies is their ability to design a strategy-making process that is capable of effectively balancing induced and autonomous strategy processes to meet the various strategic dynamics situations that their companies unavoidably face as they evolve (Burgelman and Grove, 2007).

The key strategic leadership imperative is also addressed by a significant body of literature focused on the concept of 'ambidextrous

organizations,' which are designed to handle both incremental and revolutionary change (e.g., O'Reilly and Tushman, 2008). The concept of ambidexterity is closely related to the framework of induced and autonomous strategy processes; but there remain two important differences. First, change in the induced strategy process, while 'incremental,' can be very large. For instance, developing a new microprocessor is incremental for Intel but involves hundreds of millions of dollars in development costs and billions in manufacturing investments. In the induced strategy process, 'incremental' simply means change that is well understood – doing more of what the company knows to do well. Change through the autonomous process, on the other hand, is 'radical' but is initially usually rather small. However, it always involves doing things that are not familiar to the company – doing what it is not sure it can do well. Second, change through the autonomous strategy process usually comes about fortuitously and unexpectedly, and senior and top management have initially no clear understanding of its strategic importance for the company and how it relates to the company's distinctive competencies. Resolving this indeterminacy, which is the key role played by the strategic context determination process, is the most difficult challenge facing autonomous strategic initiatives.

The analysis of the consequences of co-evolutionary lock-in confirms the importance of strategic context determination processes as a key component of the autonomous strategy process. Strategic context determination processes complement a company's structural context in important ways. They offer the possibility to suspend the selective effects of the structural context, which almost unavoidably tends to become fine-tuned for supporting the induced strategy process, and serve to create linkages between autonomous action and the company's current concept of strategy. They are the crucial nexus between exploration and exploitation and key to making the effective balancing of induced and autonomous strategy processes possible. Strategy-making as adaptive organizational capability thus depends critically on top management's capacity to allow activation and completion of strategic context determination processes in the organization's internal selection environment. Co-evolutionary lock-in, however, tends to emaciate the capacity to activate strategic context determination processes and thereby prevents a company from exploiting potentially viable autonomous initiatives. In the case of Intel, in spite of Andy Grove's efforts to vectorize everybody in the same direction, numerous autonomous strategic initiatives continued to emerge, indicating continued

attempts at exploration, but the emaciation of the company's capacity to activate strategic context led eventually to what Craig Barrett, Andy Grove's successor as Intel's CEO, called the 'creosote bush effect': Intel's core microprocessor business had begun to resemble a creosote bush, a desert plant that poisons the ground around it, preventing other plants from growing nearby. Raising awareness of this dangerous consequence may help top management manage co-evolutionary lock-in rather than being managed by it.

At a broader theoretical level, the substantive theory of co-evolutionary lock-in and its strategic consequences presented in this paper provides a potentially useful building block for further efforts to develop a formal theory of path dependence at the organizational level of analysis. General concepts such as path dependence typically provide limitedly useful guidance for field researchers struggling to identify and conceptualize novel substantive phenomena pertaining to highly complex systems. At best, such guidance may lead to finding additional illustrations of the general phenomenon as it is already understood; but more likely, it runs the risk of directing attention away from potentially interesting but less immediately identifiable novel substantive phenomena. The longitudinal research of Intel's evolution, for instance, was not guided by general theory about path dependence; yet, with the help of grounded theorizing (Glaser and Strauss, 1967) it was able to identify and conceptualize the phenomenon of co-evolutionary lock-in which, as this paper suggests, can be fruitfully related to path dependence theory. Hence, it would seem that examining – from the inside out, so to speak – the potential links between a newly discovered substantive phenomenon and more general theoretical concepts may be most likely to significantly deepen our understanding.

Finally, the analysis of the phenomenon of co-evolutionary lock-in and its strategic consequences suggests the existence of 'self-similarity across scale' (Gaddis, 2002) in various types of processes with characteristics of path dependence at different levels of social systems. Most research of path dependence, so far, has been carried out at the level of competing technologies, and at higher levels of systems (institutional, industry and organizational field). The findings of the present paper suggest that related processes may play out at the organizational level of analysis. As these findings could be conceptualized in terms of the framework of induced and autonomous strategy processes, it may turn out that this framework could also be useful for examining these processes across multiple levels of analysis.

References

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

Allison, G. and P. Zelikow (1990) *Essence of Decision: Explaining the Cuban Missile Crisis*, 2nd edn. New York: Addison Wesley Longman.

Anderson, P. (1999) 'Complexity theory and organization science', *Organization Science*, vol. 10, pp. 216–232.

Arthur, B.W. (1987) 'Competing technologies: An overview', in G. Dosi (ed.) *Technical Change and Economic Theory*. New York: Columbia University Press, pp. 590–607.

Arthur, B.W. (1989) 'Competing technologies, increasing returns, and lock-in by historical events', *Economic Journal*, vol. 99, pp. 116–131.

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

Audia, P.G., E.A. Locke and K.G. Smith (2000) 'The paradox of success: An archival and a laboratory study of strategic persistence following radical environmental change', *Academy of Management Journal*, vol. 43, pp. 837–853.

Axelrod, R. and M.D. Cohen (2000) *Harnessing Complexity: Organizational Implications of a Scientific Frontier*. London: Routledge.

Barnett, W.P. (1997) 'The dynamics of competitive intensity', *Administrative Science Quarterly*, vol. 42, pp. 128–160.

Barnett, W.P. and M.T. Hansen (1996) 'The red queen in organizational evolution', *Strategic Management Journal*, Summer Special Issue 17, pp. 139–157.

Baum, J.A.C. and B. McKelvey (eds) (1999) *Variations in Organization Science: In Honor of Donald T. Campbell*. Thousand Oaks, CA: Sage Publications.

Brown, S.L. and K.M. Eisenhardt (1997) 'The art of continuous change: Linking complexity theory and time-paced evolution in relentlessly shifting organizations', *Administrative Science Quarterly*, vol. 42, pp. 1–34.

Burgelman, R.A. (1983a) 'A model of the interaction of strategic behavior, corporate context and the concept of strategy', *Academy of Management Review*, vol. 8, pp. 61–70.

Burgelman, R.A. (1983b) 'Corporate entrepreneurship and strategic management: Insights from a process study', *Management Science*, vol. 29, pp. 1349–1364.

Burgelman, R.A. (1991) 'Intraorganizational ecology of strategy making and organizational adaptation: Theory and field research', *Organization Science*, vol. 2, pp. 239–262.

Burgelman, R.A. (1994) 'Fading memories: A process theory of strategic business exit in dynamic environments', *Administrative Science Quarterly*, vol. 39, pp. 24–56.

Burgelman, R.A. (2002a) 'Strategy as vector and the inertia of co-evolutionary lock-in', *Administrative Science Quarterly*, vol. 47, pp. 325–357.

Burgelman, R.A. (2002b) *Strategy is Destiny: How Strategy-Making Shapes a Company's Future*. New York: Free Press.

Burgelman, R.A. and A.S. Grove (2007) 'Let chaos reign, then rein in chaos – repeatedly: Managing strategic dynamics for corporate longevity', *Strategic Management Journal*, vol. 28, pp. 965–979.

Burgelman, R.A. and R.E. Siegel (2008) 'Cutting the strategy diamond in high-technology ventures', *California Management Review*, Spring, pp. 140–167.

Campbell, D.T. (1960) 'Blind variation and selective retention in creative thought as in other knowledge processes', *Psychological Review*, vol. 67, pp. 380–400.
Collins, J.C. (2001) *Good to Great*. New York: HarperBusiness.
David, P. (1985) 'Clio and the economics of QWERTY', *The American Economic Review,* vol. 75, pp. 332–337.
David, P. (2007) 'Path dependence, its critics and the quest for "historical economics" ', in G.M. Hodgson (ed.) *The Evolution of Economic Institutions: A Critical Reader.* Cheltenham, U.K. & Northampton, MA: Edward Elgar, pp. 120–142.
Farrell, J., M.K. Hunter and G. Saloner (1998) 'The vertical organization of industry: Systems competition versus component competition', *Journal of Economics & Management Strategy,* vol. 7. pp. 143–182.
Gaddis, J.L. (2002) *The Landscape of History: How Historians Map the Past.* New York: Oxford University Press.
Gersick, C.J.G. (1994) 'Pacing strategic change: The case of a new venture', *Academy of Management Journal*, vol. 37, pp. 9–45.
Glaser, B.G. and A.L. Strauss (1967) *The Discovery of Grounded Theory.* Chicago: Aldine.
Gould, S.J. (2002) *The Structure of Evolutionary Theory.* Cambridge, MA: Harvard University Press.
Grove, A.S. (1993) 'How Intel makes spending pay off', *Fortune*, February 22.
Grove, A.S. (1996) *Only the Paranoid Survive*. New York: Double Day.
Hannan, M.T. and J. Freeman (1977) 'The population ecology of organizations', *American Journal of Sociology*, vol. 83, pp. 929–984.
Hannan, M.T. and J. Freeman (1984) 'Structural inertia and organizational change', *American Sociological Review*, vol. 49, pp. 149–164.
Kauffman, S.A. (1993) *The Origins of Order: Self-Organization and Selection in Evolution.* Oxford: Oxford University Press.
Levinthal, D.A. (1997) 'Adaptation on rugged landscapes', *Management Science*, vol. 43, pp. 934–950.
Lovas, B. and S. Ghoshal (2000) 'Strategy as guided evolution', *Strategic Management Journal*, vol. 21, pp. 875–896.
March, J.G. (1991) 'Exploration and exploitation in organizational learning', *Organization Science*, vol. 1, pp. 71–87.
Meyer, A.D., V. Gaba and K.A. Colwell (2005) 'Organizing far from equilibrium: Nonlinear change in organizational fields', *Organization Science*, vol. 16, pp. 456–473.
McKelvey, B. (1997) 'Quasi-natural organization science', *Organization Science*, vol. 8, pp. 352–380.
Miller, D. (1994) 'What happens after success: The perils of excellence', *Journal of Management Studies*, vol. 31, pp. 85–102.
Miller, D. (1999) 'Selection processes inside organizations: The self-reinforcing consequences of success', in J.A.C. Baum and B. McKelvey (eds) *Variations in Organization Science: In Honor of Donald T. Campbell*. Thousand Oaks, CA: Sage Publications, pp. 93–109.
Nicolis, G. and I. Prigogine (1989) *Exploring Complexity: An Introduction*. New York: W.H. Freeman.
North, D.C. (1990) *Institutions, Institutional Change and Economic Performance*. Cambridge: Cambridge University Press.

O'Reilly, C.A. and M.L. Tushman (2008) 'Ambidexterity as a dynamic capability: Resolving the innovator's dilemma', *Research in Organizational Behavior*, vol. 28, pp. 185–206.
Pfeffer, J. and G.R. Salancik (1978) *The External Control of Organizations*. New York: Harper & Row.
Prigogine, I. (1980) *From Being to Becoming: Time and Complexity in the Physical Sciences*. New York: W.H. Freeman.
Prigogine, I. (1996) *The End of Certainty: Time, Chaos, and the New Laws of Nature*. New York: The Free Press.
Rotemberg, J.J. and G. Saloner (1994) 'The benefits of narrow business strategies', *American Economic Review*, vol. 84, pp. 1330–1349.
Rotemberg, J.J. and G. Saloner (2000) 'Visionaries, managers, and strategic direction', *Rand Journal of Economics*, vol. 31, pp. 693–716.
Sorenson, O. (2000) 'Letting the market work for you: An evolutionary perspective on product strategy', *Strategic Management Journal*, vol. 21, pp. 577–592.
Sorenson, O. and T. Stuart (2000) 'Aging, obsolescence and organizational innovation', *Administrative Science Quarterly,* vol. 45, pp. 81–112.
Thietart, R.A. and B. Forgues (1995) 'Chaos theory and organization', *Organization Science*, vol. 6, pp. 19–31.
Tsoukas, H. and R. Chia (2002) 'On organizational becoming: Rethinking organizational change', *Organization Science*, vol. 13, pp. 567–582.
Zuckerman, E.W. (2000) 'Focusing the corporate product: Securities analysts and de-diversification', *Administrative Science Quarterly*, vol. 45, pp. 591–619.

Index