

# **The Foundations of Non-Equilibrium Economics**

The principle of circular and  
cumulative causation

*Edited by*  
**Sebastian Berger**



Routledge Advances in Heterodox Economics

# The Foundations of Non-Equilibrium Economics

This thought-provoking volume presents essays on the foundations of non-equilibrium economics, i.e. the principle of circular cumulative causation (CCC). This work presents empirical research on how the interplay of technology's increasing returns to scale, institutions, resources, and economic policy leads to virtuous circles of economic growth and development, but also to vicious circles of social and ecological degradation. In particular, evidence is provided for the important role of the "development state" and strategic trade policy, economies of large-scale production in manufacturing, the regional level of development and community-based resource management regimes. While demonstrating CCC's strength in generating empirical research, the book also provides insights into its philosophical foundations and intellectual history. Several essays trace the roots of this full-fledged theoretical framework back to Adam Smith, Classical Political Economy, Thorstein Veblen, Gunnar Myrdal, K. William Kapp and Nicholas Kaldor.

As the most comprehensive collection of the growing body of CCC research to date, this book also reflects the emergence of an economic paradigm for understanding economic dynamics and for crafting viable development strategies for the 21st century. The volume will be of great interest to scholars of growth and development economics, institutional and evolutionary economics, political economy, and Post Keynesian economics from undergraduate to postgraduate research levels.

**Sebastian Berger** is Assistant Professor at Roanoke College and was awarded the 2008 Helen Potter Award by the Association for Social Services.

# **The Foundations of Non-Equilibrium Economics**

The principle of circular and cumulative causation

**Edited by Sebastian Berger**

First published 2009

by Routledge

2 Park Square, Milton Park, Abingdon, Oxon OX14 4RN

Simultaneously published in the USA and Canada

by Routledge

270 Madison Ave, New York, NY 10016

*Routledge is an imprint of the Taylor & Francis Group, an informa business*

This edition published in the Taylor & Francis e-Library, 2009.

To purchase your own copy of this or any of Taylor & Francis or Routledge's collection of thousands of eBooks please go to [www.eBookstore.tandf.co.uk](http://www.eBookstore.tandf.co.uk).

© 2009 selection and editorial matter Sebastian Berger; individual chapters the contributors

All rights reserved. No part of this book may be reprinted or reproduced or utilized in any form or by any electronic, mechanical, or other means, now known or hereafter invented, including photocopying and recording, or in any information storage or retrieval system, without permission in writing from the publishers.

*British Library Cataloguing in Publication Data*

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

*Library of Congress Cataloging in Publication Data*

The foundations of non-equilibrium economics: the principle of circular and cumulative causation/edited by Sebastian Berger.

p. cm.

Includes bibliographical references and index.

1. Equilibrium (Economics) 2. Multiplier (Economics) 3. Causation. I. Berger, Sebastian.

HB145.F68 2009

330.1—dc22

2009003520

ISBN 0-203-87373-4 Master e-book ISBN

ISBN10: 0-415-77780-1 (hbk)

ISBN10: 0-203-87373-3 (ebk)

ISBN13: 978-0-415-77780-3 (hbk)

ISBN13: 978-0-203-87373-1 (ebk)

# Contents

<i>Notes on contributors</i>	ix
<i>Foreword</i>	xii
<i>Acknowledgements</i>	xiv
<b>1 Introduction</b>	<b>1</b>
SEBASTIAN BERGER	
<b>2 On competing views of the importance of increasing returns, cumulative causation and path-dependence</b>	<b>12</b>
JOHN McCOMBIE AND MARK ROBERTS	
<b>3 Cumulative causation and Northeast Asian post-war industry policy</b>	<b>43</b>
PHILLIP TONER AND GAVAN BUTLER	
<b>4 Cumulative causation and industrial development: the regional stage</b>	<b>65</b>
GEORGE ARGYROUS AND GEOFF BAMBERRY	
<b>5 Nicholas Kaldor and cumulative causation: public policy implications</b>	<b>77</b>
RICHARD P. F. HOLT AND STEVEN PRESSMAN	
<b>6 The principle of circular and cumulative causation: Myrdal, Kaldor and contemporary heterodox political economy</b>	<b>91</b>
PHILLIP ANTHONY O'HARA	
<b>7 Circular cumulative causation à la Myrdal and Kapp</b>	<b>106</b>
SEBASTIAN BERGER	

<b>8 Utilizing the social fabric matrix to articulate circular and cumulative causation for conceptual conclusions</b>	119
F. GREGORY HAYDEN	
<b>9 Unnatural depletion and artificial abundance: a circular cumulative causation analysis of salmon fisheries and some implications for political ecological economics</b>	130
SEBASTIAN BERGER AND JAMES EDWARD GLAVIN IV	
<b>10 Circular and cumulative causation in the classics: anticipations, family resemblances, and the influence on Post Keynesian economics</b>	154
MATHEW FORSTATER AND MICHAEL J. MURRAY	
<b>11 Peirce, Veblen, and the introduction of cumulative causation into economic science</b>	172
JOHN HALL AND OLIVER WHYBROW	
<b>12 Veblen's cumulative causation and the origins of money in Mesopotamia</b>	178
ALLA SEMENOVA	
<i>Index</i>	195

# Contributors

**George Argyrous** is Senior Lecturer in the School of Social Sciences and International Relations, University of New South Wales. He has published on a range of topics, including industrial development and technological change, the history of economic thought, and the interaction between the labour market and welfare policy.

**Geoff Bamberry**, BA Dip Public Admin (QLD) Dip Teach (ASOPA) MA (Sussex) PhD (NSW), is Associate Professor of Public Administration in the School of Commerce at Charles Sturt University, Wagga Wagga, New South Wales. His research and publications have been mainly in the area of regional development, including factors influencing the clustering of manufacturing enterprises, and the development of export markets by regional firms in the food and beverage processing sector, particularly the wine industry.

**Sebastian Berger** is Assistant Professor at Roanoke College. Berger holds a doctoral degree in economics (Bremen University) and was awarded the 2008 Helen Potter Award by the Association for Social Economics. His dissertation analyses the role of Circular Cumulative Causation within Evolutionary Economics and was published by Peter Lang Verlag. His work in ecological institutional economics and theories of social costs has been published in the *Journal of Economic Issues*, *Journal of Ecological Economics*, and *Review of Social Economy*.

**Gavan Butler** received an Agr Sc and MAgSc from the University of Melbourne and a PhD from the University of California, Berkeley. Butler is an Honorary Associate of the Department of Political Economy, University of Sydney. His research interests and publications are in the areas of non-neoclassical economics, economic development, the formation of the state and public economic policy, higher education, Thai politics, and political economy.

**Mathew Forstater** is Associate Professor of Economics and Black Studies, University of Missouri – Kansas City. The Dutch and Estonian translations of his *Little Book of Big Ideas: Economics* (2007) have recently been published. Forstater has eleven entries on topics ranging from “Pan-Africanism” and “moral sentiments” to “beauty contest metaphor” and “stationary state”

in the *International Encyclopedia of the Social Sciences*, 2nd edn (2008). He is co-editor of a recent (2008) special issue of the *Journal of Black Studies* on economics and African American studies.

**James Edward Glavin IV** holds a Bachelor of Arts in Environmental Science, with a focus on Ecological and Energy Issues, from the University of Missouri – Kansas City. He is the founder and president of the Energy Awareness Group of UMKC, a Peak Oil advocacy and education student group. James also served as the Vice-President of the Environmental Studies Student Association and the Campus Green Party and has been featured as a community leader and a guest expert at numerous speaking engagements.

**John Hall** completed his PhD at the Graduate Faculty of the New School for Social Research in 1985, the same year that he joined the Economics Faculty of Portland State University. Each academic year, Hall lectures in Institutional Economics and History of Economic Thought.

**F. Gregory Hayden**, Professor of Economics at the University of Nebraska-Lincoln, has devoted his research to three areas since he received his doctorate in economics at the University of Texas-Austin. They are the development of (1) the social fabric matrix which is a methodology for modelling complex systems in real-world contexts, (2) a transdisciplinary meta-policy-making paradigm, and (3) the application of instrumental philosophy to policy analysis. The three are integrated in his book, *Policymaking for a Good Society* (Springer 2006).

**Richard P. F. Holt** is Professor of Economics at Southern Oregon University. He has authored and edited a number of books and articles including the prize-winning *The Changing Face of Economics: Conversations with Cutting Edge Economists*, with David C. Colander and Barkley Rosser, Jr. (2004). He has published with Steven Pressman *A New Guide to Post Keynesian Economics* (2001), and *Empirical Post Keynesian Economics* (2006). They are currently working on the third volume, *Post Keynesian Economics and Public Policy* (forthcoming).

**John McCombie** is Director of the Cambridge Centre for Economic and Public Policy, Department of Land Economy, University of Cambridge, and Fellow in Economics, Downing College Cambridge, UK. He is currently an editor of *Spatial Economic Analysis* and was an editor of *Regional Studies*. His research interests include the study of national and regional growth disparities, EU regional policy, economic growth and the balance-of-payments constraint, and criticisms of the aggregate production function and conventional measures of technical progress.

**Michael J. Murray** is Assistant Professor of Economics, Central College, Pella Iowa. He is also a PhD candidate in Economics and Mathematics at the University of Missouri-Kansas City and a Research Assistant for the Center for Full Employment and Price Stability.



**Phillip Anthony O'Hara** is Professor of Global Political Economy and Governance in the Global Political Economy Research Unit (GPERU) and the Department of Economics, Curtin University. He has published five books and won the Gunnar Myrdal Prize from the European Association for Evolutionary Political Economy, and the Clarence Ayres Award from the Association for Evolutionary Economics. He specializes in the principles of political economy, post-neoliberal governance, and global political economy.

**Steven Pressman** is Professor of Economics and Finance at Monmouth University in West Long Branch, NJ. He also serves as co-editor of the *Review of Political Economy*, and as Associate Editor and Book Review Editor of the *Eastern Economic Journal*. He has published around 100 articles in refereed journals and as book chapters, and has authored or edited ten books, including *A New Guide to Post Keynesian Economics* (Routledge, 2001) and *50 Major Economists*, 2nd edn (Routledge, 2006).

**Mark Roberts** is lecturer in Spatial Economics, Planning and Policy at the University of Cambridge, and Fellow of Murray Edwards College, Cambridge. His work includes applied spatial econometric analysis of (sub-)regional productivity and growth differences in the UK and Europe. He does research with the World Bank's Development Economics Research Group (Sustainable Rural and Urban Development Unit), and contributed to the 2009 World Development Report. He is also co-editor of the journal *Spatial Economic Analysis*.

**Alla Semenova** is a PhD candidate in Economics and Social Science at the University of Missouri – Kansas City. She is also an adjunct Professor of Economics at Roanoke College, and a research associate with the Center for Full Employment and Price Stability. In her dissertation, she explores the questions of money's origins and early development. Her research on the origins of money in ancient Mesopotamia was recognized with the outstanding graduate student paper award by the Association for Institutional Thought in 2009.

**Phillip Toner** is Senior Research Fellow in the Centre for Industry and Innovation Studies at the University of Western Sydney. His PhD thesis was published as *Main Currents in Cumulative Causation: The Dynamics of Growth and Development* (1999). His research areas include industrial structure analysis and policy; the economics of technical innovation, and comparative national skills formation systems. He has undertaken research for the OECD, Industry Canada, the Australian Government, and industry associations and unions.

**Oliver Whybrow** graduated with honours from Portland State University with a Bachelor of Science in Economics. After working with Stifel Nicolaus in Baltimore, Maryland, he attended the University of Leuven in Leuven, Belgium where he is currently working towards a Master's in International Business Economics. Oliver is published in the *Journal of Economic Issues* and is a copy editor for *Panoeconomicus*, the *Serbian Journal of Economic Sciences*. He is a member of the American Economic Association.

# Foreword

Geoffrey M. Hodgson

“Do not adjust your theory – reality is at fault.” This could be the slogan of much of mainstream economics since the Second World War. The slogan fits because since the rise of neoclassical theory at the end of the nineteenth century, mainstream economics has regarded the determination of equilibrium conditions as the Holy Grail of theoretical discovery. But in order to demonstrate the existence of equilibria within models, economists have typically had to assume diminishing returns and negative feedback. Once we enter a real world with increasing returns and positive feedback – a world where deviations can be amplified rather than suppressed – then the conventional demonstrations of equilibria are no longer viable.

A major theme in the history of modern economics is the attempts of a minority within the profession to remind the equilibrium theorists of the importance of positive feedback mechanisms, even before that term was invented by Norbert Wiener in 1948. Alfred Marshall noted in Appendix H of his *Principles* (1890) that increasing returns could undermine the conditions for an equilibrium of supply and demand. In his *Interest and Prices* (1898) Knut Wicksell wrote of a “cumulative process” of interaction between prices, the rate of interest and investment. Wicksell influenced fellow Swede Gunnar Myrdal, who formulated a model of cumulative causation in his *Monetary Equilibrium* (1931) and used the core idea in his later studies of racial discrimination, uneven regional growth and underdevelopment. Previously Allyn Young published a seminal article in the *Economic Journal* in 1928, emphasizing that economic change “propagates itself in a cumulative way”. In turn, Young taught Cambridge economist Nicholas Kaldor, who was a staunch critic of equilibrium economics and also influenced by Myrdal. Later W. Brian Arthur had to remind the profession of the importance of positive feedback in a series of articles dating from the 1980s. Nobel Laureate Paul Krugman has also written on these themes, but only with limited acknowledgement of the pioneers in this area.

The term “cumulative causation” dates from Thorstein Veblen’s famous article “Why is economics not an evolutionary science?”, published in the *Quarterly Journal of Economics* in 1898. But he used it in a different way. Instead of positive feedback, Veblen used “cumulative causation” to describe an extended sequence of causal links, without beginning or end. Inspired by Darwinism, he

understood that phenomena could not be adequately explained in terms of their presumed purposes or destinations. Explanation had to be in terms of the causal sequence, showing how each stage led to the next. But (again without using the term explicitly) there are cases where Veblen discusses processes of positive feedback, and he was highly critical of equilibrium approaches.

Young was one of Veblen's admirers, and they were both together in Stanford University in the early 1900s. Young eventually moved to the London School of Economics in the 1920s, where he met an untimely death from pneumonia. Yet he is a key link between the institutionalism of Veblen and European Keynesians such as Kaldor and Myrdal.

Modern economic systems contain multiple processes among heterogeneous agents with positive and negative feedbacks. Consider the dynamics of boom and bust. Just as a boom in stocks or house prices encourages more buyers, who push up prices further, a downturn encourages selling, which drags down prices still more. These are processes of positive feedback. But eventually negative feedback kicks in. Some investors observing the protracted boom may become wary that it may end, and some observing a slump may perceive an opportunity to buy bargains. This counter-cyclical behaviour may eventually become more widespread, overcome the positive feedback and turn the market around. Other examples of negative feedback are the operation of "automatic stabilisers" such as lower taxes and higher unemployment benefits as the economy enters a recession. These income enhancements increase effective demand and help to counter the downward forces.

Instead of being driven by the search for equilibria within models, which leads to the rejection of positive feedbacks that make the mathematical search more complicated or even intractable, economists should start from the real world. The relative importance of positive or negative feedbacks cannot be settled a priori. It is an empirical matter. But it is necessarily aided by heuristic models of the type established by some of the aforementioned authors. Reality should drive the theory – not the other way round.

Such a realist spirit pervades the present volume. It is a highly appropriate reminder not only of the importance of cumulative causation, but also that economists are under an obligation to understand the messy world around them, rather than to retreat into the aesthetic technicalities of their models. I welcome the chapters here as further contributions to ongoing research in this area.

# Acknowledgements

The editor is grateful to the *Journal of Economic Issues* for permission to reprint papers and versions of papers that were previously published. In addition, I would like to thank Rolf Steppacher from the K. William and Lore L. Kapp Foundation for the Humanization of the Social Sciences for access to Kapp's unpublished manuscripts and helpful discussions on Kapp's approach to economics. I wish to thank Edward Nik-Khah and John Hall for helpful comments on the subject of this book. For help with the manuscript I am grateful to Dreama Poore and Anastasia Juers. For institutional support I am indebted to Roanoke College.

# 1 Introduction

*Sebastian Berger*

This introduction provides the reader with an explanation for publishing a book on theories of circular cumulative causation (CCC) at this moment in time. In other words, it explains how the CCC theories offered, for example, by Nicholas Kaldor, Gunnar Myrdal, Thorstein Veblen and K. William Kapp are relevant for economic science. Consider, for instance, that the Bank of Sweden Prize in Economics in Memory of Alfred Nobel (2008) was awarded to Paul Krugman for his contributions to economic geography and trade theory. Both contributions include increasing returns as a key concept (Prize Committee of the Royal Swedish Academy of Sciences 2008: 3), that has been an integral part of CCC theories. Krugman's prize-winning core-periphery model works with a mechanism of self-reinforcing causation to explain migration from agricultural to industrialized regions and, thus, reflects what Gunnar Myrdal – himself a winner of the Bank of Sweden Prize – had discussed much earlier in his analysis of circular cumulative causation:

The model is driven by the location choices of firms and individuals. [...] there is an element of circular causality [...] [setting] in motion a cumulative process [...] Krugman was able to build a strict model of the process of circular causation discussed much earlier by Myrdal (1957).

(ibid.: 14)

In addition, there are further significant developments within economics that attest to CCC's increasing importance. The major revitalization of non-equilibrium economics, for instance, demonstrates the growing popularity of approaches that can grasp the real dynamic and self-reinforcing aspects of economic phenomena. In this, CCC has been acknowledged as a key concept of evolutionary-institutional economics (Berger 2007) and a "common denominator" concept for many non-equilibrium research areas (O'Hara 2007). The important signal coming from these major developments is that CCC theories and their intellectual traditions have become increasingly important. Given this situation, one would expect to find a well-established consensus in the literature on what CCC exactly is and what it can do.

Indeed a good way to start is to present the existing consensus of CCC theories. One of the defining characteristics of all CCC theories is that they fruitfully

capture important dynamic aspects of economic reality that are not reflected by the mechanistic metaphor of a (stable) equilibrium in the neoclassical standard model. Such aspects include economic growth, technological change, business cycles, socio-economic and ecological change and so on. CCC theorists, furthermore, reject the abstract formalism of the neoclassical method for its lack of empirical grounding. Consequently, CCC approaches also discard the *a priori* notion of optimal economic outcomes that are inherent in the neoclassical standard model. Instead, several CCC theories emphasize that such notions are the result of implicit value judgements made by neoclassical researchers that have to be exposed and opened up for discussion. Several currents in CCC theory exhibit a strong empirical research interest in economic disparities, social costs and economic crises that are not perceived as minor accidental deviations from the “normal” optimal case but as major and systemic patterns worthy of research in their own right.

However, a glance at the relevant literature also raises several fundamental questions: What makes CCC a principle rather than a theory, a hypothesis, a concept, a methodology or even a paradigm? If there are different meanings of CCC, are these already fully understood and their potentials for economics fully exploited? What are the methodological and normative implications of different CCC approaches and are they compatible? A survey of CCC theories makes clear that a consensus on these questions has yet to emerge. This is largely due to the fact that there are different currents in CCC theory with unique perspectives that lead to different answers regarding the questions above. Of course, this diversity has been the source of CCC’s large body of fruitful research. Yet this diversity of perspectives has also been the reason for the lack of a more unified and perhaps more influential approach. One underlying cause may be that not all economists agree with the political economy of Myrdal, Veblen and Kapp that is intertwined with their CCC theories. Thus, their influence seems confined to certain groups of, for example, evolutionary-institutional and ecological economists but also to other disciplines, such as sociology. Whatever the case may be, the increasing interest in CCC theories demands a response to the existing open questions and perhaps an answer as to what the next step in the development of CCC could be. Taking stock of the status quo of CCC theories in the context of current developments in economics contributes to such a clarification and possibly also to building a framework for analysis that coherently integrates the diverse CCC currents.

Consequently, this book has set itself two main goals. First, by presenting new research on its diverse intellectual origins, as well as new applications of CCC, it brings diverse currents of CCC approaches into conversation with one another. The book provides a comprehensive account of CCC’s origin, philosophical foundations, applications, and implications for economic theory. Several essays point out the differences as well as similarities within the different strands of CCC theory. Second, and closely interrelated with the first goal, the volume aims to promote the use of CCC in economic analysis by demonstrating the fruitfulness of CCC research. The reader is provided with a collection of essays covering research areas such as growth and development economics, economic policy, ecological economics, economic geography, trade

theory, classical political economy, Post Keynesian economics as well as evolutionary-institutional economics. The chapters are arranged in such a way that broadly speaking the Kaldorian current is presented first, leading up to the tradition of Myrdal and Kapp, which is finally followed by classical political economy and Veblenian contributions. The following introduction highlights some of the implications of the research presented in this book.

## Continuity, openness and self-reinforcing causation

Chapter 2 (McCombie and Roberts) focuses on the cumulative causation of increasing returns as one of the main sources of economic growth. In addition to providing empirical evidence for the existence of returns to scale, the authors introduce the reader to the important debate between Walrasian general equilibrium theorists and Kaldor's theory of increasing returns. The chapter also includes a comparison with more recent approaches, such as Krugman's new economic geography and path-dependence. This evidences CCC's enormous potential for developments in economics surrounding self-reinforcing causation. Indeed, a look at recent developments in economics shows that self-reinforcing dynamics are the focus of many approaches, such as self-organization (Foster 2005; Witt 1997), system dynamics (Radzicki 2003) path dependence (Arthur 1994) and evolutionary game theory (Gintis 2000), and are often referred to as "non-linearities" or "positive feedbacks".

Chapter 10 (Forstater and Murray) links CCC to Post Keynesian contributions by Nell and Passinetti. Yet, further links to post-Keynesian economics can be pointed out, namely Minsky's financial instability hypothesis that embodies self-reinforcing causation of expectations in economic boom and decline (Minsky [1978] 1985: 37–8, 45–6). This research theme can be traced back to Keynes's *General Theory* (1936) but also to Myrdal's earlier cumulative causation theory in *Monetary Equilibrium* (1933), which underlined the crucial role of expectations in macro-economic instability, preceding Keynes's theory in important respects. Kaldor admitted that it was his reading of the German version of Myrdal's work that made him an easy convert to Keynes's general theory three years later (Barber 2008: 27, 30). Minsky was also an institutional economist, so it is not surprising that Veblen's *Theory of Business Enterprise* (Veblen 1904: 113) had already described accounting of business capital in an expansion as a self-reinforcing inflationary "system of make-believe" that gives rise to a further extension of credit with the purpose of further expanding production and sales. This leads to the interesting question as to what makes CCC unique other than being the first concept to capture self-reinforcing causation for socio-economic analysis (Myrdal 1944; Richardson 1991).

Several chapters in this volume provide answers to what makes CCC so unique. They may be grouped under the notions of openness or continuity (Hall and Whybrow, Chapter 11). In essence, this means the systematic incorporation of a broader set of factors that have to be considered as endogenous (Forstater and Murray, Chapter 10) because of their circular interdependency with the open

economic system (Berger, Chapter 7): for example, the state (Toner and Butler, Chapter 3), socio-cultural factors (O'Hara, Chapter 6), ecological variables (Berger and Glavin, Chapter 9), history and the substantive economy (Semenova, Chapter 12). These complex interactions also require tools for policymaking (Hayden, Chapter 8; Holt and Pressman, Chapter 5) that do not only look at market interactions. Thus, CCC is crucially important because it is very flexible and more generally applicable than other approaches in evolutionary economics. The importance of the notion of openness for evolutionary economics has been defended by Geoffrey Hodgson in a recent debate against Ulrich Witt's narrowly defined endogenous market causation (Hodgson 2004: 365). It is noteworthy that Myrdal's contribution to CCC theory was awarded the Bank of Sweden Prize (jointly with Friedrich von Hayek) for the "penetrating analysis of the interdependence of economic, social and institutional phenomena" and for successfully carrying out interdisciplinary research (Sandelin 1991: 216; Barber 2008: 164–7). CCC's holistic view truly improves the quality of economic research and offers a unique potential for analysing self-reinforcing causation that goes far beyond narrow technology adoption in markets. The notion of circular causation economics "comes to be characterized [...] as an inquiry into the bearing which all facts have upon men's economic activity" (Veblen 1900: 262); or, in cumulative causation terminology:

[An] inquiry into the cultural or institutional development as affected by economic exigencies or by the economic interests of the men whose activities are analyzed and portrayed [...] a cumulatively unfolding process or an institutional adaptation to cumulatively unfolding exigencies.

(Veblen 1900: 263–4)

## **Economic growth and development: trends and the taboo of teleology**

Chapter 4 (Argyrous and Bambergy) focuses on stages of growth related to increasing returns. The authors provide empirical evidence for the existence of stages in industrial development that are, however, not inevitable. The authors build on Kaldor's contribution that considered stages as junctures where the self-reinforcing virtuous circle could break down unless government policy was favourable to the transition. This approach avoids much of the teleological implications that were the reason for Myrdal's roundabout rejection of standard stage theories and their conservative political implications.

Rejecting teleology and finding ways to conceptualize trends of change is a central concern within evolutionary economics, and CCC's self-reinforcing causation also implies the notion of a trend. The notion of a trend is, for example, one of the reasons for Hodgson's critique of Marxian economics and his alternative of universal Darwinism. The latter seems to be inspired by Darwinian evolutionary biology and its principle of "undirected" biological evolution (Mongioli 2008). So, should economics reject trends along with teleology based on Darwinian evolutionary biology? It seems noteworthy that even in the evolu-



tion of biological systems trends exist, and there is a variety of scientific hypotheses to explain them. In addition, there is something unique about the principle underlying social dynamics because the societal level of organization is not identical to the natural.

Against this background the reader will find it of interest that Myrdal did not resort to the notion of evolution or Darwin but instead used the term “dynamics”. The term “dynamic” was, according to Tillich (1933), traditionally used by progressive political orientations and revolutionary romanticism. It denotes being that is in the movement from its potential to its reality/actuality, or a being that is not yet completely formed, but embodies the potential and the power of a form. The term “dynamic” is often misused to denote the opposite of “static” or “resting”, thus destroying its original meaning. Its origin is “dynamis”, i.e. a productive potentiality that urges to its own actualization. This is the meaning that Aristotle attached to the term and “dynamis” denotes a system imbued with an inherent propelling force. This fits into the Aristotelian tradition of explaining events in terms of the actualization of inherent powers by the triggering action of external circumstances.

However, modern physics and Darwinian biology usually impose a taboo against teleology and final causality. In economics, the taboo of teleology was most prominently introduced by Veblen’s concept of “blind cumulative causation”. Despite their non-teleological character biological and social researchers have had to deal with evidence of existing real tendencies (Fernández 2008: 6). This is where CCC offers a way to conceptualize trends for the purpose of social inquiry. Self-reinforcing causation may be used as a tool for building hypotheses about dynamics:

Where Darwin’s theory of natural selection is based on the principle of evolution, the theory of human development, which presupposes Darwin’s theory, is based on the vicious-circle principle. And where the principle of evolution came to constitute the core of biology, the vicious-circle principle is intended to constitute the core of human ecology.

(Dilworth 2002: 78)

CCC can serve as a hypothesis about trends that are individuated by spatio-temporal circumstances and that do not last forever. The vicious or virtuous circles embodied in CCC theory are not considered inevitable and the hypothetical character of the CCC approach prevents dogmatic teleology. Importantly, CCC does not aim at establishing a specific kind of causation (e.g. self-reinforcement) as the normal case of all systems and does not aim at a delimitation of a range of facts via taxonomy or uniformities.

## **Values and the trend to naturalize the social sciences**

Toner and Butler’s (Chapter 3) research results highlight the role of the state and economic policy for initiating and furthering trends of industrialized growth.

Their chapter implies that the discussion is brought back to Myrdal and Kapp who both considered the state (i.e. democratic public action) as crucial for a virtuous circle of development. This raises important normative questions: What role should industrial growth have in the development effort? What kind of industrial growth should be pursued? Who benefits and what are the limits and side-effects of growth? What constitutes wealth? Addressing these questions in the tradition of Adam Smith's "Wealth of Nations", Myrdal and Kapp pointed to the importance of value judgements and substantive (normative) rationality in the development process, particularly with regard to economic disparities and ecological disruption that persist despite and due to industrial growth. Since the late 1950s Kapp contributed to what became the so-called "eco-development" movement in the 1970s, nowadays known as "sustainable development". In this the evaluation of cumulative changes in the quality of life is holistic and normative, encompassing social costs as important wealth-diminishing aspects of growth and development. And Myrdal chose "The Equality Issue in World Development" as a topic for his Nobel Memorial Prize speech. In this he argued for countering the self-reinforcing cumulative increase of wealth disparities by redistributing resources from the rich industrialized countries to the poor countries mainly via "a much more frugal life style so far as growth in consumption, and production for home consumption, of many material products is concerned" (Myrdal in Barber 2008: 166). Thus, normative considerations are at the core of CCC theories.

Myrdal's CCC operates with a "normative" research hypothesis of a vicious circle, i.e. "social waste" and inefficiencies as a result of self-reinforcing causation (Berger 2007). This approach makes CCC a veritable alternative to recent developments to "naturalize" economics either by verbal or formal analogies to processes that take place on the organic level of organization, e.g. self-organization (Foster 2005; Witt 1997), Universal Darwinism (Hodgson 2002), evolutionary game theory (Gintis 2000) and genetic algorithms (Axelrod 1997). Essentially these approaches argue that ontologically the notions of non-equilibrium, self-organization and evolution span the social and the natural level. This allegedly makes social theory compatible with insights from the natural sciences, i.e. evolutionary biology (Darwin) and physico-chemistry (Prigogine). Mirowski identifies this as the "separate but equal doctrine" as one out of four states of minds on the natural-social relationship (Mirowski 1994: 12). Several economists consider this trend to "naturalize" economics via formal models, verbal analogies or metaphors derived from the natural sciences as a "multilayered power game" in the "furtherance of particular human interests", i.e. shoring up legitimacy, trying to disenfranchise political economists (Mirowski 1994: 13). The application of natural science analogies and ontologies in economics often obscures underlying value premises and purposes, i.e. the political elements that guided the choice of the analogy (Geisendorf 2001; Kubon-Gilke 1996; Vromen 1997). As Myrdal emphasized throughout his work, the social scientist cannot escape the political element so that value premises have to be made explicit to avoid implicit and hidden manipulation

(Myrdal 1929). Also, consider that Kapp argued that applying CCC as a hypothesis about increasing economic disparities and ecological disruption satisfies the conditions of scientific method, as defined by John Dewey's instrumentalism, because it grows out of actual social tensions or needs that are related to ends-in-view, i.e. a plan or policy for the resolution of the conflicting situation (Dewey 1938: 499). In conclusion the advantage of CCC is that it does not resort to the natural sciences to analyse social phenomena and that it does not avoid explicit reference to ends and values, i.e. the political element in the analysis of social causation.

In addition, there is another fundamental problem with the trend to naturalize economics. Even critics from within evolutionary economics have recently recognized that social organization and its mechanisms of change are far more complex and not constant through time so that analogies taken from natural sciences contribute practically no additional insights into socio-economic relations (Nelson 2001, Rosenberg 1994). Sceptics also argue that analogies from natural sciences lead to futile "checklist approaches", i.e. the search for similarities between the organic and the socio-economic units of analysis and mechanisms of change (Vromen 2004). Kapp saw this danger of reasoning by analogies as early as 1961, arguing that reasoning by analogies makes it possible to dispense with the need to formulate clear notions of the characteristics of the social units of analysis. By imposing analogical reasoning upon the material studied, the collections of data, testing and so on tend to lose their specificity. In addition, events which are not captured by the analogy may even be neglected and withdrawn from investigation: "Once the intellectual operation based upon the analogy is in full swing, it is usually too late to remind oneself of the imperfect character of the original analogy upon which the whole enterprise rests" (Kapp 1961: 58). Instead, Kapp applied Myrdal's CCC to analyse the dynamic interrelation between humans and society in his important book *Towards a Science of Man in Society* (1961). The advantage of CCC is that it offers a way to analyse dynamic social phenomena without prematurely resorting to analogies taken from natural sciences where there is no urgent necessity to do so.

This is not to say, however, that loose heuristic metaphors cannot be useful as a first step in the creative stage of associative thinking and understanding. Mirowski is right in pointing to the healthy side effects from "spiral narratives" that bring nature and society into interplay (Mirowski 1994: 15–16). This concerns, for example, the application of CCC in the context of biotic resources (salmon fisheries) (see Berger and Glavin, Chapter 9, this volume). This application is not to be understood as a reversed "separate but equal doctrine" that tries to anthropomorphize nature but as an attempt to illustrate CCC's strength in holistic causal analysis for understanding the complex dynamics of natural resources, technology and economic institutions. While it can be useful to point out instances of circular cumulative causation in the natural as well as in the social systems, CCC theorists realize that the higher level of complexity of human society is not governed by purely functional relationships but is subject to volition and deliberation. Nevertheless, CCC's focus on continuousness (see

Hall and Whybrow, Chapter 11, this volume) points to the importance of interdisciplinary knowledge for understanding, for instance, the dynamic interrelations between the biological, psychological and cultural structure of the human organism and its influence on economic development. Another example are the laws of thermodynamics and biological open system's theory that help us understand the material-energetic level of the economic process which is part of the larger bio-geo-chemical level of analysis.

### **CCC and economics in the “calculable future”**

This introduction has discussed CCC's place in the context of current developments in economics, pointing out the concept's advantages over other approaches in evolutionary economics. Now the interesting question remains as to what are the chances that CCC becomes part of “the teaching of economics to prepare economists of the future for the tasks with which they will be increasingly concerned” (Kapp 1976: 103). The answer to such a question is necessarily speculative but Veblen (1925) e.g. predicted that in the “calculable future” “loosely speaking, no argument on economic matters will get a reasonable wide hearing until it is set out as a ‘business proposition’ in terms drawn from the conduct of business administration, business finance, national trade, salesmanship and publicity” (Veblen 1925: 53). Veblen's prediction seemed correct when half a century later Kaldor recognized that

equilibrium theory has reached the stage where the pure theorist has successfully demonstrated that the main implications of this theory cannot possibly hold in reality, but has not yet managed to pass this message down the line to the textbook writer and to the classroom.

(Kaldor [1972] 1978: 180)

At about the same time Kapp presumed that

[i]t is possible that the desire to retain the traditional doctrine may [...] give rise to a “conceptual freeze” [and] it is unlikely that this freeze will be broken in the calculable future under the impact of new facts, new evidence of environmental disruption, new catastrophes and an increasing public opposition to the deterioration of the physical and social environment.

(Kapp 1976: 105)

Myrdal, on the other hand, predicted optimistically that

within the next ten or twenty years [...] [a] more institutional approach will win ground, simply because it is needed to deal in an effective way with the practical and political problems that now tower over us, and which threaten to overwhelm us.

(Myrdal 1976: 86)

While the development in economics since the 1990s was doubtlessly characterized by a revitalization of evolutionary-institutional economics as well as an increasing pluralism in economics, neoclassical economics with its teleological notions of equilibrium and optimality remains the dominant approach, in particular, in the undergraduate curriculum. In addition, approaches that provide “naturalizing” evolutionary explanations to economic phenomena seem to have become more popular than “political” CCC approaches. Despite of Myrdal’s above prediction failing to come to pass, this book takes up his inspiring optimism. It aims at revitalizing CCC for the analysis and resolution of real world economic problems by providing a comprehensive demonstration of CCC’s capacity to produce insightful research.

## References

- Arthur, B. W. (1994) *Increasing Returns and Path Dependence in the Economy*, Ann Arbor: University of Michigan Press.
- Axelrod, R. (1997) *The Complexity of Cooperation: Agent-based Models of Competition and Collaboration*, Princeton, NJ: Princeton University Press.
- Barber, William J. (2008) *Gunnar Myrdal – An Intellectual Biography*, New York: Palgrave Macmillan.
- Berger, Sebastian (2007) *Europäischer Institutionalismus – Die Kernkonzepte Open System Approach und Circular Cumulative Causation und ihr Beitrag zur modernen evolutischen institutionellen Ökonomie – Unter besonderer Berücksichtigung der Beiträge von Nicholas Georgescu-Roegen, K. William Kapp, Adolph Lowe, Gunnar Myrdal, and Karl Polanyi*, Frankfurt am Main: Peter Lang Verlag (Dissertation: Monography).
- Cohen, I. B. (1994) *Interactions: Some Contacts between the Natural Sciences and the Social Sciences*, Cambridge, MA: MIT Press.
- David, P. A. (2002) “Path Dependence, its Critics, and the Quest for Historical Economics”, in T. Cowen and E. Crampton (eds) *Market Failure or Success: The New Debate*, Cheltenham (UK) and Northampton, MA: Edward Elgar.
- Dewey, J. (1938) *Logic – The Theory of Inquiry*, New York: Henry Holt and Company.
- Dilworth, C. (2002) “Population, Technology, and Development – The Vicious-circle Principle and the Theory of Human Development”, in M. N. Dobkowski and I. Wallimann (eds) *On the Edge of Scarcity – Environment, Resources, Population, Sustainability, and Conflict*, Syracuse, NY: Syracuse University Press.
- Fernández, E. (2008) “Peircian Realism, Critical Realism, and Some Epistemic Problems of the Social Sciences”, paper presented at University of Missouri – Kansas City Seminar.
- Foster, J. (2001) “Competition, Competitive Selection and Economic Evolution”, in P. Garrouste and S. Ioannides (eds) *Evolution and Path Dependence in Economic Ideas: Past and Present*, Northampton, MA: Edward Elgar.
- (2005) “The Self-organizational Perspective”, in K. Dopfer (ed.) *The Evolutionary Foundations of Economics*, Cambridge: Cambridge University Press.
- Geisendorf, S. (2001) *Evolutionäre Ökologische Ökonomie*, Marburg: Metropolis.
- Gintis, H. (2000) *Game Theory Evolving: A Problem-centered Introduction to Modeling Strategic Interaction*, Princeton, NJ: Princeton University Press.
- Hodgson, G. M. (2002) “Darwinism in Economics: From Analogy to Ontology”, *Journal of Evolutionary Economics*, 12, 3: 259–81.

- (2004) “Book Review. Essays on the Evolutionary Approach to Economics, by U. Witt”, *Journal of Evolutionary Economics*, 14: 365–8.
- Kaldor, N. [1972] (1978) “The Irrelevance of Equilibrium Economics”, in N. Kaldor (ed.) *Further Essays in Economics*, London: Holmes & Meier Publishers.
- Kapp, K. W. (1961) *Toward A Science Of Man in Society – A Positive Approach to the Integration of Social Knowledge*, The Hague: Martinus Nijhoff.
- (1976) “The Open-system Character of the Economy and Its Implications”, in K. Dopfer (ed.) *Economics in the Future*, London: Macmillan.
- Kubon-Gilke, G. (1996) “Institutional Economics and the Evolutionary Metaphor”, *Journal of Institutional and Theoretical Economics*, 152, 4: 723–38.
- Minsky, H. P. [1978] (1985) “The Financial Instability Hypothesis: A Restatement”, in P. Arestis and T. Skouras (eds) *The Post-Keynesian Economic Theory: A Challenge to Neo-Classical Economics*, Sussex: Wheatsheaf Books; Armonk, NY: M. E. Sharpe.
- Mirowski, P. (1994) “Doing What Comes Naturally: Four Metanarratives on What Metaphors Are For”, in P. Mirowski (ed.) *Natural Images in Economic Thought: “Markets Read in Tooth and Claw”*, Cambridge and New York: Cambridge University Press.
- Mongiovi, G. (2008) “Bookreview. Economics in the Shadow of Darwin and Marx: Essays on Institutional and Evolutionary Themes by G. Hodgson”, *EH.NET*, pp. 6. Available online [en.net/bookreviews/library](http://en.net/bookreviews/library) (accessed March 2008).
- Myrdal, G. ([1929] 1954) *The Political Element in the Development of Economic Theory*, Cambridge, MA: Harvard University Press.
- (1944) *American Dilemma: The Negro Problem and Modern Democracy*, New York and London: Harper & Row.
- (1976) “The Meaning and Validity of Institutional Economics”, in K. Dopfer (ed.) *Economics in the Future*, London: Macmillan.
- Nelson, R. (2001) “A Viewpoint on Evolutionary Economic Theory”, in Y. Aruka (ed.) *Evolutionary Controversies in Economics: A New Transdisciplinary Approach*, Tokyo and New York: Springer.
- O’Hara, P. A. (2007) “Principles of Institutional-evolutionary Political Economy – Converging Themes from the Schools of Heterodoxy”, *Journal of Economic Issues*, 41, 1: 1–42.
- Prize Committee of the Royal Swedish Academy of Sciences (2008) “Scientific Background on the Sveriges Riksbank Prize in Economic Science in Memory of Alfred Nobel 2008 – Trade and Geography – Economies of Scale, Differentiated Products and Transport Costs”: 1–23.
- Radzicki, M. J. (2003) “Mr. Hamilton, Mr. Forrester, and a Foundation for Evolutionary Economics”, *Journal of Economic Issues*, 37, 1: 133–73.
- Richardson, G. P. (1991) *Feedback Thought in Social Science and Systems Theory*, Philadelphia, PA: University of Philadelphia Press.
- Rosenberg, A. (1994) “Does Evolutionary Theory Give Comfort or Inspiration to Economics”, in P. Mirowski (ed.) *Natural Images in Economic Thought: “Markets Read in Tooth and Claw”*, Cambridge and New York: Cambridge University Press.
- Sandelin, B. (1991) “Beyond the Stockholm School”, in B. Sandelin (ed.) *The History of Swedish Economic Thought*, London and New York: Routledge.
- Skott, P. (1994) “Cumulative Causation”, in G. M. Hodgson, W. J. Samuel and M. R. Tool (eds) *The Elgar Companion to Institutional and Evolutionary Economics* (Vol. 1), Aldershot: Edward Elgar.
- Tillich, P. (1933) *Die Sozialistisch Entscheidung*, Potsdam: Alfred Protte Verlag.

- Veblen, T. (1900) "The Preconceptions of Economic Science", *The Quarterly Journal of Economics*, 14, 2: 240–69.
- (1904) *The Theory of Business Enterprise*, New York: Charles Scribner's Sons.
- (1925) "Economic Theory in the Calculable Future", *American Economic Review*, 15, 1: 48–55.
- Vromen, J. (1997) "Evolutionary Economics: Precursors, Paradigmatic Propositions, Puzzles and Prospects", in J. Reijnders (ed.) *Economics and Evolution*, Cheltenham: Edward Elgar.
- (2004) "Conjectural Revisionary Economic Ontology: Outline of an Ambitious Research Agenda for Evolutionary Economics", *Journal of Economic Methodology*, 11, 2: 213–47.
- Witt, U. (1997) "Self-organization and Economics – What is New?", *Structural Change and Economic Dynamics*, 8, 4: 489–507.
- (2005) "The Evolutionary Perspective and The Firm", in K. Dopfer (ed.) *The Evolutionary Foundations of Economics*, Cambridge: Cambridge University Press.

## 2 On competing views of the importance of increasing returns, cumulative causation and path-dependence

*John McCombie and Mark Roberts*

### Introduction

Nearly two centuries ago, the ratio of GDP per capita in the richest to the poorest region of the world was about 3:1, and this difference could be largely attributed to variations in land fertility and other geographic advantages.<sup>1</sup> Today, however, the ratio exceeds 70:1. A key question is: What has caused this dramatic divergence? One possible answer lies in a rapid and sustained increase in the rate of technical change in the now advanced countries. However, such an answer, in turn, begs its own question, namely: Why did this acceleration occur in the first place?

The most plausible answer dates back to Adam Smith's *The Wealth of Nations*, written on the eve of Britain's industrial revolution in 1776. It involves the dramatic impact that the division of labor, or increasing returns to scale broadly defined, has in raising productivity. Book I of *The Wealth of Nations* begins with the famous dictum, "The greatest improvement in the productive powers of labor, and the greater part of the skill, dexterity, and judgment with which it is any where directed, or applied, seem to have been the effects of the division of labor." Smith then proceeded to elaborate on the remarkable increase in productivity that specialization brought in the case of a "very trifling manufacture," namely the trade of a pin-maker.<sup>2</sup> He noted that the way this very simple object was manufactured was, in fact, highly specialized.

One man draws out the wire, another straightens it, a third cuts it, a fourth points it, a fifth grinds it at the top for receiving the head; to make the head requires two or three distinct operations; to put it on is a peculiar business, to whiten the pins is another; it is even a trade by itself to put them into the paper; and the important business of making a pin is, in this manner, divided into about eighteen distinct operations, which in some manufactories are all performed by distinct hands, though in others the same man will sometimes perform two or three of them.

By this means the workers could, through specialization, produce over 240 times the amount of pins that they could manufacture if each worker dealt with the



production from start to finish. This is undoubtedly the most famous example of internal increasing returns. However, Smith realized that the returns stretched beyond the individual firm:

The separation of different trades and employments from one another seems to have taken place in consequence of this advantage. This separation, too, is generally carried furthest in those countries which enjoy the highest degree of industry and improvement; what is the work of one man in a rude state of society being generally that of several in an improved one.

Indeed, it would be no exaggeration to argue that much of economic growth from the earliest times has been associated with the increasing division of labor. Per capita growth first started with animal husbandry and crop cultivation that produced an agricultural surplus. This surplus, in turn, permitted the development of a flourishing artisan class by medieval times. But per capita growth really took off with the industrial revolution in England, and this was associated with an acceleration of the division of labor. As Kaldor (1977) perspicaciously noted, while major inventions undoubtedly played a role in eighteenth-century industrialization, none of these would have had the same impact had it not been for the development of the factory system. The rapid gains in productivity through increasing returns, broadly defined, that this system permitted, led to profits being ploughed back into industry, thereby facilitating rapid accumulation. This, in turn, enabled capital-intensive inventions (such as Watt's steam-engine, the use of coal, rather than charcoal, to smelt iron and the "spinning jenny") to be introduced.<sup>3</sup>

Kaldor (1977: 196), in a sentence that could have been written by Marx, wrote: "since increasing returns (the economies of large-scale production) appeared virtually inexhaustible, each entrepreneur strove to accumulate capital as fast as possible in order to keep ahead of, or at least keep pace with, his rivals." Today, the degree of specialization in production is more than anything Smith could have imagined. It does not require much of a thought experiment to appreciate how low the level of productivity would be if everybody had to be self-sufficient, producing everything personally, even with free access to current technology, capital and raw materials.

But one of the great conundrums in the history of economic thought is the marginalization of the concept of increasing returns in mainstream, or neoclassical, economics. Indeed, ever since Ricardo's (1817) *Principles*, the alternative assumption of constant returns has held sway: most notably through the progressive elaboration and formalization of competitive general equilibrium, the concept of Pareto optimality and the two fundamental theorems of welfare economics. While increasing returns has not been totally ignored, it has generally been treated in an ad hoc manner. That is until the past 20 years or so, when the pendulum has swung almost to the other extreme with the development, in turn, of strategic trade theory, endogenous growth theory and the new economic geography.<sup>4</sup>

Kaldor (1972: 378) argued, with some justification, that economic theory went wrong in the middle of the fourth chapter of Book I of *The Wealth of Nations*. Here Smith discusses the theory of value and implicitly assumes constant returns to scale. Ricardo was the first to develop a formal economic deductive model and he was more concerned with diminishing marginal returns. This was largely due to the rising price of corn at the time and the fixity of agricultural land. This, he predicted, would lead to a falling agricultural profit rate, which, in turn, would result in the dismal stationary state where, in Ricardo's words, "profits are so low as not to afford [the capitalists] an adequate compensation for their trouble and the risk which they must necessarily encounter in employing their capital productively."

Diminishing marginal returns are not necessarily incompatible with increasing returns, but Ricardo developed the marginal principle with respect to the determination of land rents.<sup>5</sup> It was this that was then generalized by the subsequent neoclassical economists into a general marginal productivity theory of distribution. The powerful application of Euler's theorem showed that if factors are paid their marginal products, the product is completely exhausted, provided there are constant returns to scale (and perfect competition). The underlying motivation could have been, for example, as Dobb (1973) held, ideological, since with the marginal productivity theory, factor rewards are determined by the technical conditions of production, thereby banishing the Marxian notion of the exploitation of labor from the stage.<sup>6</sup> Even if there were not these ideological undertones, however, the great advantage was that the marginal productivity theory provided a simple way of closing the neoclassical model and solving the "problem" of distribution. This could be accomplished by simply invoking the first order conditions of the production function, together with the conditions for perfect competition. Such analysis could, furthermore, be seamlessly transplanted into macroeconomics by ignoring the insuperable and serious aggregation problems (Fisher 1992) through the fiction of the "representative firm."

There are, of course, other more plausible explanations for the determination of factor shares. Suppose, for example, that firms pursue a mark-up pricing policy (Lee 1998). In this case, if  $\pi$  is the size of the mark-up, then labor's share is equal to  $1/(1 + \pi)$ . The parameter  $\pi$  is determined by the degree of competition in the industry<sup>7</sup> and the state of the economy as a whole, which influences the relative bargaining power of labor and capital. This approach can easily encompass increasing returns to scale without any internal inconsistency, but it is more difficult to close the standard neoclassical deductive model with such a specification.

Of course, the "problem" of increasing returns did, on occasion, attract attention. It was used to justify protectionist policies in the late eighteenth and early nineteenth centuries. Alexander Hamilton, the first US Secretary to the Treasury, advanced the argument in 1794, albeit unsuccessfully, that the US should adopt tariffs to protect its fledgling industries from competition from the UK, thereby advancing the "infant industry" argument. He was directly influenced by Adam Smith's observations concerning the division of labor and the importance of

increasing returns to scale (but not his views on free trade). Fredrick List (1841) was, in turn, influenced by Hamilton and subsequently made a similar argument in favor of protectionism in the early stages of industrialization.<sup>8</sup>

Alfred Marshall, in particular, was fully aware of the possible, even likely, existence of downward-sloping supply curves and multiple equilibria, as well as of the irreversibility of increasing returns. The latter introduces historical time and path-dependence into the picture. But these qualifications to the perfectly competitive model were relegated to Appendix H of the *Principles* (Marshall 1890) and never became central to his analysis. As Stigler (1951) points out, Adam Smith, in fact, had created a dilemma of which Marshall was well aware. As increasing returns lower costs as production expands, the eventual outcome should be monopoly, yet some competitive industries existed. But to assume perfectly competitive markets meant that Smith's dictum was empirically insignificant, which was implausible. As value theory moved to bring microeconomic production and consumer theory to the fore, and with the former, the neoclassical theory of the firm, Marshall (1890) tried to resolve this dilemma in two ways.

First, he introduced the concept of external economies of scale. Thus, a firm increasing its output would not find its costs falling as a result: this would only occur as the industry as a whole expanded, thereby allowing the concept of perfect competition to be maintained.<sup>9</sup> This was never wholly convincing at the time, as it was difficult to find convincing examples of external economies of scale, apart from the benefits of a common pool of skilled labor or a trade journal.<sup>10</sup> Nevertheless, as noted above, Smith had himself discussed the importance of the division of labor increasing between trades, a point echoed by Allyn Young (1928) and Kaldor (1972) (although neither considered that it was compatible with perfectly competitive markets).

Second, Marshall postulated the decline of entrepreneurial ability as a particular firm grew, which offset the gains from internal increasing returns. However, Marshall's biological analogy (joint stock companies, although subject to increasing internal returns, are like the trees in the forest: they grow and then stagnate, but "do not readily die") is not compelling. His desire to preserve the deterministic results of perfect competition meant that, in practice, he relegated all increasing returns to the category of an externality.

The implications of increasing returns also arose in the cost controversies of the 1930s, with Sraffa (1926) pointing out the logical problems increasing returns pose for the competitive model. Clapham (1922) injected some empirical questions in a largely theoretical debate, to which we return below. Robinson (1933) and Chamberlin (1933) subsequently introduced the concept of monopolistic competition, which allows for production under conditions of internal increasing returns. However, this approach never really took off until the pioneering paper by Dixit and Stiglitz (1977) demonstrated how it might be modeled in a tractable mathematical manner.

Arrow (2000: 171) has presented another reason for the persistence of the convexity assumption, arguing that the "history of competitive equilibrium is

essentially *cumulative*” (emphasis in original). In other words, “each step depends in a fairly clear cut way on the work of predecessors.” The development occurred in a linear fashion from Adam Smith (at least from Chapter IV, Book One of *The Wealth of Nations*), through Ricardo, Menger, Jevons, Walras to Hicks, Koopmans, Arrow and Debreu. “Other aspects of perfectly competitive theory, such as the theory of demand, optimality theorems, and the relation between competitive equilibrium and bargaining outcomes, also have well-structured histories in which earlier work influenced the later work which subsumed it” (Arrow 2000, p. 172). The paradigm was logically self-consistent and, at the theoretical level, was replete with Kuhnian “puzzles” to solve.

In analyzing growth, Solow (1956) and Swan (1956) developed their well-known growth model based on constant returns to scale. By suggesting that the growth of factor inputs was relatively small in accounting for the growth of productivity, Solow (1957) placed the emphasis on the rate of technical change. Such technical change was initially treated as exogenous, since there was no plausible formal economic theory of technical progress at the time.<sup>11</sup> Although growth theory did begin to give serious consideration to the concept of increasing returns from the early 1980s with the emergence of endogenous growth theory,<sup>12</sup> the Solow–Swan model was to subsequently reassert itself with the empirical work of Mankiw, Romer and Weil (1992) and controversy surrounding both the knife-edge properties and counterfactual predictions concerning scale effects of the early *AK* endogenous growth models. Although more recent models reconcile these problems with endogenous growth, in doing so, they adopt many of the features of the traditional Solow–Swan model, with its emphasis on constant returns (Jones 2001).

### **Young, Myrdal and Kaldor on increasing returns and Walrasian general equilibrium theory**

While, historically, the attitude of the mainstream economics profession towards the concept of increasing returns has been, at best, lukewarm, it is nevertheless possible to find a continuum of economic thought emphasizing the importance of increasing returns. Veblen (1915) sought to explain why Germany’s level of development overtook that of the UK in spite of the latter’s early start. He attributed this to the fact that the UK was locked into obsolete technology, especially with respect to the railway system. Because of technical interrelatedness, new investment was added piecemeal to the obsolete, but nevertheless still functioning, capital equipment.<sup>13</sup> For greatest efficiency the whole system would have had to be replaced simultaneously, but the improved economies would not have compensated for this. Hence, the problem for the UK lay in the indivisibility of the transport network, non-convexity and sunk costs.

Increasing returns brings with it path-dependence where “history matters.” Small events in the past, arising by chance, may have large and disproportionate effects on later developments. Under these circumstances it is impossible to analyze a sequence of economic events without recourse to historical, as opposed

to logical, time. Comparative statics, for example, is of little use in understanding the process of change. As Arthur (1989) pointed out, a major cause of path-dependence is the existence of multiple (stable) equilibria. Under these conditions, small “chance” events<sup>14</sup> can determine in which equilibrium state the system ends up. And a major cause of multiple equilibria is increasing returns to scale.

It was Young (1928) who first fully realized the important implications of Adam Smith’s division of labor in a path-breaking, yet somewhat idiosyncratic, paper. As we shall see, this idea was later taken up and extended by his former pupil Kaldor, who argued that it undermined most of neoclassical general equilibrium theory. Rosenstein-Rodan (1943) stressed the role of indivisibilities and, hence, increasing returns in economic development. Myrdal (1957) developed the notion of “circular and cumulative causation” while, at about the same time, Hirschman (1958) criticized the notion of a balanced growth path. Both Myrdal and Hirschman stressed how growth was initially spatially polarized and only later gave rise to spread or backwash effects to the surrounding areas. Since the related approaches of all these economists have been analyzed in detail by Toner (1999), in this section we shall concentrate on Kaldor’s challenge to general equilibrium theory, and Hahn’s (1973) subsequent riposte. We shall return to the issue of path-dependence in the next section.

Kaldor enunciated his critique of “equilibrium economics” in several public lectures between 1972 and 1984. This critique grew out of the dissatisfaction that Kaldor felt with the then prevailing state of economic theory, dissatisfaction that Kaldor claimed he shared with other contemporary leading lights in the British economics academy.<sup>15</sup> Given the time he was writing, Kaldor normally associated “equilibrium economics” with Walrasian general equilibrium theory as elaborated by, *inter alios*, Arrow, Debreu and MacKenzie. However, this was not always the case, and Kaldor was quite explicit in the fact that he saw his critique as extending to “dynamic” general equilibrium theory in the guise of neoclassical (Solovian) growth theory (Kaldor 1975: 348; 1979).<sup>16</sup> More generally, it seems reasonable to portray Kaldor’s critique of “equilibrium economics” as taking place on two levels. First, we not only have Kaldor’s specific criticisms of certain of the key axioms and points of emphasis of Walrasian general equilibrium theory, but second, his more general methodological criticisms. The latter may be taken as applying not only to general equilibrium theorists, but also more widely to the practices of mainstream economists. Not only this, but at least some of his criticisms of Walrasian general equilibrium theory may be interpreted as representing a specific application of his more general methodological approach.

Kaldor most fully articulated his critique in his 1972 paper. Drawing heavily on Young’s (1928) paper, he emphasized Adam Smith’s dictum that the “division of labor is determined by the extent of the market.” This, as we have seen, is far from a tautology. It implies that as the volume of production grows, so the degree of specialization both within and between firms increases, leading to an increase in productivity, which, in turn, further extends the market. An implication of this is that the capital intensity of production is more a function of the

scale of production than of relative factor prices.<sup>17</sup> Here, we have the foundations of Kaldor's cumulative causation model (Kaldor 1970).<sup>18</sup> Kaldor further stressed the role of "learning by doing" (Arrow 1962) or "dynamic" economies of scale, the gains from which are irreversible. The end result is that economic progress is *path-dependent*.

Once, however, we allow for increasing returns, the forces making for continuous change are *endogenous* – "they are engendered from within the economic system"<sup>19</sup> – and the actual state of the economy during any one "period" cannot be predicted except as a result of the sequence of events in previous periods which led up to it. As Young put it, with increasing returns "change becomes progressive and propagates itself in a cumulative way."<sup>20</sup> Further, "no analysis of the forces making for economic equilibrium, forces which you might say are tangential at any moment of time, will serve to illumine this field, for movements away from equilibrium, departures from previous trends, are characteristic of it".<sup>21</sup>

(Kaldor 1972: 1244; emphasis in original)<sup>22</sup>

At the general methodological level, Kaldor considered the practices of mainstream economists to be unscientific. This conclusion was based upon his definition of science as "a body of theorems based on assumptions that are *empirically* derived (from observations) and which embody hypotheses that are capable of verification with regards to the assumptions and the predictions" (Kaldor 1972: 1237; emphasis in original). As such, the axioms or basic assumptions of "scientific theory" are "chosen on the basis of direct observation of the phenomena *the behaviour of which forms the subject-matter of the theory*" (Kaldor 1972: 1238; emphasis added). In contrast, Kaldor identified the axioms or basic assumptions typically made by neoclassical economic theorists as being either unverifiable (e.g. constrained optimization of objective functions) or directly contradicted by observation (e.g. the existence of perfect foresight) (Kaldor 1972: 1238; 1979).

Kaldor further pointed out that where economic theory does make reference to concepts that, *prima facie*, seem to have counterparts in reality, upon closer inspection any similarity is purely superficial. This is a consequence of the fact that no attempt is made to show how such axiomatic concepts as goods and processes of production "are to be defined or recognized *in relation to empirical material*" (Kaldor 1972: 1238; emphasis added; see also 1979, 1996). As a symptom of this "pre-scientific" state of development, Kaldor identified the tendency for economic theory to experience fads and fashions such as the rise and fall of the Philips curve or, one might add, rational expectations in the guise of the new classical microeconomics (Kaldor 1972: 1240). Nor did he see econometrics, at least as practiced by the mainstream, as providing the missing link to reality: "where empirical material is brought into conjunction with a theoretical model, as in econometrics, the role of empirical estimation is to 'illustrate' or to 'decorate' the theory, not to provide support to the basic hypothesis" (Kaldor 1972: 1239, fn. 1).



Progressing to his specific criticisms of Walrasian general equilibrium theory, Kaldor directed his attack along two main lines. First, he criticized the theory for overly concentrating on the *allocative* functions of markets to the neglect of their *creative* functions (Kaldor 1972: 1240; see also 1985: 14). As a corollary of this, he argued that such theory tends to emphasize issues of substitutability when it should be emphasizing “the essential complementarity between different factors of production [...] or different types of activities [...] which is far more important for an understanding of the laws of change and development of the economy” (Kaldor 1975: 348).

Most notably in this respect, Kaldor criticized mainstream theory for ignoring the essential complementarity between broad sectors of the economy such as agriculture and mining on the one hand and manufacturing on the other (see, *inter alios*, Kaldor 1979). Furthermore, applying his general methodological criticisms, Kaldor took Walrasian general equilibrium theory to task for its specific axiom of convexity. For him, this was an unscientific assumption that he felt should be replaced by the assumption of increasing returns (Kaldor 1972: 1241–6). He based this argument on direct observation of the historical growth and development of capitalist economies in general and Britain in particular (Kaldor 1977).<sup>23</sup>

However, it is important to note that Kaldor’s agenda was not purely nihilistic. Thus, at the methodological level, he advocated a “stylized-facts” or inductive approach to theory building (Kaldor 1985: ch. 1). In contrast to the axiomatic-deductive practices of mainstream economics, this approach involves, first of all, identifying empirical regularities whether through the use of quantitative or qualitative methods. The aim is then to “seek the most reasonable explanation capable of accounting for these ‘facts’, independently of whether they fit into the general framework of received theory or not” (Kaldor 1985: 8). Crucially, Kaldor saw these stylized facts not as strict empirical regularities that admit no exception, but as only being “true in the broad majority of observed cases – in a sufficient number of cases to call for an explanation that would account for them” (Kaldor 1985: 9). Furthermore, Kaldor conceived of any explanation of a stylized fact as necessarily being both partial and fallible. Thus, hypotheses that are formulated to explain stylized facts “relate to particular *aspects* of the economy and they may be suggestive of others. They may be discarded if they prove inconsistent with other observed features and then be replaced by something else” (Kaldor 1985: 9; emphasis in original). Meanwhile, in the more specific case of Walrasian general equilibrium theory, he saw the axioms and basic assumptions of the approach as entailing an *ahistoric* view of the world.

More specifically, he saw Walrasian equilibria as necessarily being *determinate* equilibria and such equilibria permit no meaningful role for history because they imply that “whatever the initial situation, the system will converge on a *unique point* the exact nature of which [...] can be deduced from the ‘data’” (Kaldor 1972: 1244; emphasis in original). However, Kaldor saw such a view of the world as contradicting his observations of economic and social reality. This is because such observations led him instead to a vision of an “Economics without Equilibrium” (Kaldor 1985) by which he meant an “Economics without *determinate*

Equilibrium.” Such an economics captures the notion that economic processes in general, and the growth and development process in particular, are path-dependent ones in which “history matters” in the most fundamental of senses.

### **Hahn’s (1973) riposte to Kaldor**

Kaldor’s (1972) paper was thus an attack on the relevance of Walrasian general equilibrium theory, and he argued that once one allows for increasing returns to scale (removes “the scaffolding”), the whole concept becomes untenable. In response, Hahn (1973), in his inaugural lecture, provided a robust defense of general equilibrium theory. While Kaldor did make some relatively minor slips in his criticisms of general equilibrium theory,<sup>24</sup> there remain two substantial issues of contention between Kaldor and Hahn. The first is methodological, while the second concerns the implications of increasing returns to scale.

With respect to the first issue, Hahn (1973: 22) argued that “Professor Kaldor believes that the received theory is vacuous by virtue of being unfalsifiable.” However, for Hahn, much of the usefulness of a rigorous general equilibrium theory is its essentially negative role. For example, if someone argues that there is no need to worry about exhaustible resources, as increasing scarcity will be captured by the price mechanism, all one needs to do is to look at the stringent conditions, as set out by the general equilibrium theorists, for this to be true to realize how implausible this assertion is. This is a defensible, albeit rather limited role for general equilibrium and it concedes that it is not a scientific theory neither in the Popperian, nor Kaldorian, sense of the term. Nevertheless, general equilibrium theory has since been found to face formidable theoretical problems that greatly undermine its usefulness even in this context (see Kirman 1989).

As far as the second issue is concerned, Hahn (1973: 12–13; emphasis added) argues: “an Arrow–Debreu equilibrium may exist when there are increasing returns. Not only is this so when these increasing returns are not internal to firms, but even if they are, *provided they are not too large*.” As this would seem to be a major criticism of Kaldor, it is worth considering the conditions under which increasing returns *do not* pose difficulties for general equilibrium theory. An early demonstration of this was the seminal paper by Farrell (1959), generalized by Starr (1969), which was cited by Hahn (1973). Farrell shows diagrammatically that for profit-maximizing firms in a competitive market, there must be a convex negatively sloped *aggregate* transformation curve in the case of two outputs,  $X$  and  $Y$ . Suppose each firm experiences increasing returns to scale so that their *individual* transformation curves are concave. Farrell (1959: 386–9) demonstrates that while a particular firm may specialize in producing either  $X$  or  $Y$ , if they have different production functions and their transformation curves are ranked by their overall slopes, the aggregate transformation curve can still take an approximately smooth, convex form. This is provided that there are a large number of firms.

Hence, “the allocative efficiency of competitive markets survives concave production functions, whether they reflect economies of specialization or of scale, so long as the output of each individual producer is negligible” (Farrell



1959: 388). However, if the number of firms is relatively small and perfect competition breaks down, then the problems emphasized by Kaldor for general equilibrium theory materialize. Moreover, as Marshall (1890) and Sraffa (1926) long ago pointed out, increasing returns is likely to lead in the long run to only a small number of firms, or, in the limiting case, a single monopolist. For Kaldor, industrial structure is highly oligopolistic and the number of firms insufficient to maintain perfect competition in Farrell's sense.

Hahn, however, denies the revolutionary implications that Kaldor sees caused by the prevalence of increasing returns to scale.

Now we say that a given path taken by the economy is production inefficient if there is an alternative one which gives us more of some good at some time and not less of any good at any time. There is nothing in the economy here discussed which makes such an ordering impossible. If we take finite time horizons, as long as we like, and suppose the set of alternatives closed, then an efficient path also exists. It is simply a muddle to go from the difficulties increasing returns impose for perfect competition to the view that allocation does not matter. Indeed, the truth is orthogonal to the view. For the more important increasing returns are, especially the dynamic variety, the greater the potential losses from misallocation.

(Hahn 1973: 31)

This is not being totally fair to Kaldor who did not say that allocation did not matter, but rather that the notion of a unique optimal allocation of resources loses all meaning, except in the short run. As we shall see below, the economy may be moving along a path that is possibly more efficient than any other path at a particular point in time, but still end up locked into a path that, from a longer run perspective, does not maximize welfare (Arthur 1988, 1989, 1994; David 1975, 1985, 1986, 1997, 1999, 2001, 2007). Moreover, if the path opens up new opportunities that, due to increasing returns, only become apparent when the economy moves along that path, there may be other *unknown* paths that would have been preferable *if* they had been apparent. In this sense, Hahn's assumption that "the set of alternatives [is] closed" is crucial because it rules out precisely such paths. Thus, whereas Hahn implicitly assumes all future states of the economy to be known, Kaldor's point is simply that we do not know what the other path is likely to be. Although Kaldor did not put it in these terms, what he seems to have had in mind is the idea that the actual path of the economy is governed by a dynamic stochastic process.

### **Path-dependence and increasing returns to scale**

This divergence of views between Kaldor and Hahn has a strong echo in the criticisms of Liebowitz and Margolis (1995) of the work of David (1985) and Arthur (1989) on path-dependence. David and Arthur argue that, even with optimizing agents, the trajectory an economy follows will be subject to path-dependence and

this means that, in all probability, it will be sub-optimal. Liebowitz and Margolis are, however, intent on showing that path-dependence has either trivial consequences or that it does not pose a major challenge to the neoclassical paradigm.<sup>25</sup>

Liebowitz and Margolis begin by identifying what they see as three categories of path-dependence. *First-degree path-dependence* is where the outcome is determined by some set of initial conditions and, while it cannot be changed without cost, the outcome is optimal. The fact that the hands of clocks move clockwise, rather than counter-clockwise, is a good example.<sup>26</sup> *Second-degree path-dependence* occurs when decisions are taken in the absence of knowledge about the alternatives; in other words, when there is imperfect foresight. "Sensitive dependence on initial conditions leads to outcomes that are regrettable and costly to change. They are not, however, inefficient in any meaningful sense, given the assumed limitations on knowledge" (Liebowitz and Margolis 1995: 207). The development of the petrol motor car engine, rather than the steam-powered alternative engine, *may* be a case in point. It is "may," because we can only speculate that the steam-engine would have been more efficient. *Third-degree path-dependence* leads to an outcome that is inefficient, but is potentially remediable. In other words, there exists some method or set of institutional arrangements by which the preferred outcome could have come about, but this was not achieved. It is only the latter that provides a challenge for the neoclassical paradigm, and indeed it can simply be subsumed under the well-known category of market failure.

The issues may be brought out clearly by reference to a simple example from Arthur (1989) that is used by Liebowitz and Margolis (1995). Arthur uses the example to show how an economy can become locked into an inefficient technology. In particular, he considers the case of two technologies, *A* and *B*, which are adopted sequentially by agents.<sup>27</sup> Over time the payoffs from both technologies increase. The payoffs, which are assumed here to be shared by both new and old adopters alike, are shown in Table 2.1. (It is thus assumed that when, say, technology *A* has been adopted by 30 agents, the payoff to all the previous adopters of this technology increases to 13.)<sup>28</sup> It may be seen that for each agent, it is rational to adopt technology *A*. (This assumes that agents attempt to maximize their immediate payoffs when they make their decision as to which technology to adopt.) This is true even though, in the long run, it would be optimal for the first adopters to accept a relatively lower payoff and invest in technology *B*. This is because after 30 adoptions, the payoff to technology *B* is greater than to *A*. It may also be shown that if there had been 60 adoptions of *B*, the cumulative shortfall of the early adopters of *B* would have been more than offset by the subsequent gains.<sup>29</sup>

Liebowitz and Margolis's point is that, if the adopters know all the payoffs, then a unique Nash equilibrium, which is also the optimum solution, exists. In other words, all agents will opt for *B*. They argue that a *sine qua non* for a sub-optimal lock-in to occur is for there to be a lack of perfect knowledge. But this is still compatible with the neoclassical model of optimizing behavior, as the optimization is conditional on a given information set and this is simply *second-degree path-dependence*. For path-dependence to pose some problem for neoclassical economics, it needs to be shown that there was some remedial

Table 2.1 Hypothetical payoffs from two competing technologies as a function of the number of previous adoptions

No. of previous adoptions	0	10	20	30	40	50	60	70	80	90	100
Period	0	1	2	3	4	5	6	7	8	9	10
Technology A	10	11	12	13	14	15	16	17	18	19	20
Technology B	4	7	10	13	16	19	22	25	28	31	34

Source: Adapted from Arthur (1989, Table 2).

action that could have been taken, but was not. For example, it is necessary to show that a subsidy providing the necessary incentives for agents early on in the adoption process to choose technology *B* rather than technology *A* could have been provided, but was not. If a tax adjustment or subsidy can affect the path, Arthur (1989) terms this process as *flexible*. However, many paths are not flexible. This is the case, for example, when the costs of a subsidy are prohibitively large, especially considering that the finance for such a subsidy may have to come from current real resources. For example, sufficient funding may not be available if payoffs for *A* and *B* are for large projects such as either the development or reconstruction of the railway network or the steel industry, rather than the oft-quoted, but comparatively minor, example of the competition between VHS and Betamax video recorders (see Cusumano *et al.* 1992).<sup>30</sup>

Furthermore, let us assume that there is a specific time dimension to the adoption of the technology and in each period, ten agents make a decision as to which technology to adopt. There is a certain rate of discount which, when applied to the stream of future earnings, makes the net present value in period 0 of technology *A* greater than that of *B*. This poses no problem for the argument of Liebowitz and Margolis who would simply argue that technology *A* is the optimal, or efficient, choice. But assume that the agents have finite lives. This results in an intergenerational problem. At some point, from the viewpoint of later generations, technology *B*, if it had been adopted from the start, would have the greater net present value.<sup>31</sup> In other words, in our example, this would occur no later than period three, no matter what the discount rate. At this point, the net present value of *B* exceeds that of *A*. Consequently, which technology is the more efficient depends upon the specific cohort of agents under consideration.

Most importantly, Liebowitz and Margolis's approach towards the concept of path-dependency is confused, as they apply it to a deterministic process. As David (2001: 21; emphasis in the original) points out:

That path dependence is a property of *stochastic* sequential processes is *not* mentioned [by Liebowitz and Margolis 1995], and only the allusion to "contingency" provides any hint of the subject's probabilistic context. [...] Even that slender clue, however, is disguised by the statements that would have us associate path dependence with *deterministic* chaos, and the property of "sensitive dependence on initial conditions" which characterizes that class of dynamic systems.

David provides two definitions of path-dependence that emphasize the stochastic nature of these systems, one negative and one positive: “*Process that are non-ergodic, and thus unable to shake free of their history, are said to yield path dependent outcomes*” and “*a path dependent stochastic process is one whose asymptotic distribution evolves as a consequence (function of) the process’s own history*” David (2001: 19; emphasis in original). The second definition implies that, as a system progresses through time, the distribution of equilibrium outcomes at any point in time is determined by the historical trajectory traversed. Moreover, this probability distribution is changing and, as a consequence, the system is always chasing a moving target that it is never able to hit.

This distinction between deterministic systems and stochastic path-dependent systems is crucial. Suppose there are multiple equilibria, and the final outcome is determined solely by the (historical) initial conditions, then the path followed is deterministic. Let us make this clearer by some examples. Suppose all the assumptions underlying the Solow (1957) neoclassical growth model are fulfilled, including that all countries have the same rate of technical progress and the same investment–output ratio. We observe two countries whose productivity growth rates differ and neither is growing at the steady-state rate. The reason for the growth rate disparities is that there is absolute convergence; with the difference in the growth rates of the two countries being due to the different initial capital–labor ratios. The two economies will thus be following different trajectories, but will be converging to the same steady–steady growth rate. This is not path-dependence as defined by David. The convergence process in the Solow model is *ergodic*. *Ergodicity* is where a different set of historical events (or initial conditions) leads to the same equilibrium outcome with probability one.

Another example is Thirlwall and Dixon’s (1979) formalization of the Kaldor cumulative-causation model, but with the inclusion of a balance-of-payments constraint. Give plausible values of the parameters, this model has a steady-state rate of growth determined by the ratio of the income elasticity of the demand for the region’s exports to its income elasticity of demand for its imports. These are exogenously given in the model, but differ empirically between countries. As in the Solow model, if for some reason the country is not at its steady-state growth rate, it will converge to the latter and the exact path will depend upon the initial conditions and the parameters of the model. But again, this is not path-dependent in David’s sense of the term as the trajectory of the economy can be predicted exactly once the initial conditions (including the parameters of the model) are known.<sup>32</sup>

Part of the confusion arises from Liebowitz and Margolis’s (1995) choice of Arthur’s (1989) example discussed above, which is deterministic. This was used by Arthur as a heuristic device merely to illustrate the concept of lock-in and not path-dependence. Arthur’s full model is crucially driven by a stochastic process. He assumes that there are two different types of agents (for convenience denoted by R and S). Whereas R initially finds that technology *A* is more profitable than technology *B*, for S the opposite is the case, but their payoffs from each technology are also contingent on the number of adopters of this technology. (This is where the increasing returns to scale enter the picture. The more a technology is

adopted, and, by implication the greater its production, the more its cost falls and the greater its payoff becomes.)<sup>33</sup> The agents sequentially adopt a technology in random order, but eventually one technology, say *A*, will get sufficiently far ahead that its payoff to both *R* and *S* exceeds the payoff from the other technology. Once this situation occurs, both types of agent will adopt the technique and the economy is locked into *A*, which will eventually capture the entire market.

There are a number of points to note about this. Although, *a priori*, we cannot tell whether *A* or *B* will eventually dominate the market, we know with probability one that one of them eventually will, even though the agents arrive at random.<sup>34</sup> By contrast, in the constant returns case, the payoffs are independent of the market shares. Thus, although there may still be uncertainty surrounding the exact market shares at any point of time, there is perfect *predictability* as to the long-term split of market shares. In this case “the small degree of uncertainty built in ‘averages away’ so that the observer has enough information to pre-determine market shares accurately in the long-run” (Arthur 1989: 118).

We are now in a position to return to, and resolve, the differences between Kaldor and Hahn. In the quotation cited above, Hahn (1973) argues that “there is nothing here in the economy discussed that makes such an ordering [of production efficient paths] impossible. If we take finite time horizons, as long as we like, and suppose the set of alternatives closed, then an efficient path exists.” This is describing a deterministic economy with multiple equilibria. Kaldor, although he does not explicitly state so, has a view of economic change as a non-ergodic process – “the actual state of the economy during any one ‘period’ cannot be predicted<sup>35</sup> except as a result of the sequence of events in the previous periods which led up to it.” It should be noted that while a stochastic process can be factored into the neoclassical concept of equilibria, it has to be of a form that “averages out.”<sup>36</sup> A good example of this is the treatment of uncertainty by assuming rational expectations in the new classical macroeconomics. Here the agents know correctly the underlying structure of the model of the economy, subject to uncertainty that produces non-systematic errors. Thus, over a period of time agents will, on average, not make mistakes. As Paul Davidson has pointed out on numerous occasions (see e.g. Davidson 1972), this is to treat uncertainty as an ergodic rather than non-ergodic phenomenon.

### **The empirical evidence – aggregate production function studies and the Verdoorn law**

It may be seen from the above that whether or not there are empirically substantial returns to scale is central to the discussion. Notwithstanding Smith’s example of the pin factory, the empirical evidence concerning increasing returns today is still regarded by some as ambiguous. As Romer (1994: 10) commented, “if you are committed to the neoclassical mode,” the data “cannot be used to make you recant. They do not compel you to give up the convenience of a model in which markets are perfect.”<sup>37</sup> Nevertheless, in this section we shall show that when a correctly specified model is estimated, there is overwhelming evidence that there are substantial returns to scale, at least in industry.

It is now over 80 years since Paul Douglas and Charles Cobb published their seminal estimation of an aggregate production function using time-series data for the US economy (Cobb and Douglas 1928). This paper first introduced the Cobb–Douglas production function, and its estimation, to the profession<sup>38</sup> and sparked off a plethora of subsequent production function studies.

Douglas was a labor economist and was motivated by the concern that his colleagues were drawing downward-sloping demand curves for labor, yet had no idea of the value of the elasticity of demand for labor. His early time-series estimations suffered from a number of specification problems. There was no allowance for technical change or changes in capacity utilization and constant returns to scale were imposed. In fact, the early studies were received with great hostility owing to multicollinearity, simultaneity and identification problems. Mendershausen (1938) was so skeptical as to be of the opinion that the results should be torn up and thrown away (see Griliches and Mairesse (1995) for a discussion of these early criticisms). The a priorists at Chicago University also believed the empirical exercise to be a waste of time. Faced with such a reception, Douglas came close to abandoning his work completely.

But he persevered. By 1940, Douglas had published, with a number of colleagues, estimates of numerous aggregate production functions with unconstrained coefficients using cross-sectional (both cross-state and cross-industry) data with, all told, hundreds of observations. While econometrics was still in its infancy and Douglas relied on OLS, he found that not only did some of the estimated coefficients differ very little from unity, but also that they were very close to their respective factor shares.<sup>39</sup> This led him to conclude: “a considerable body of independent work tends to corroborate the original Cobb–Douglas formula, but more important, the approximate coincidence of the estimated coefficients with the actual shares received also strengthens the competitive theory of distribution and disproves the Marxian” (Douglas 1976: 914). By the time Douglas delivered his presidential address to the American Economic Association in 1948, the growing econometric evidence overwhelmingly suggested the presence of constant returns to scale.

Subsequent cross-sectional estimates using US state data by Hildebrand and Liu (1957) and Moroney (1972) found broadly comparable results, as did Klotz (1970) using data for 17 four-digit manufacturing industries, and Griliches and Ringstad (1971) using Norwegian cross-firm data.

All of these studies reported close statistical fits of the Cobb–Douglas production function with the data, with  $R^2$ s around 0.9 not uncommon. There have been numerous subsequent studies of aggregate production functions over the years employing increasingly more sophisticated econometric techniques. Although there is not the space to survey these studies here, it is sufficient to say that time-series estimates generally do not give such good fits as cross-section data for the same reasons Douglas found all those years ago.

A major problem for those who find plausible the existence of large increasing returns to scale at the economy-wide level is the so-called *scale problem*. If there are increasing returns to scale at the national level, then large countries should always have higher productivity levels than smaller countries, even if

they share the same level of technology. For example, how can we explain the fact that Switzerland (population 7.5 million) has a productivity level approximately the same as the US (population 301 million)? Some idea of the orders of magnitude involved can be gained by considering a simple thought experiment. In particular, consider a hypothetical country, country *A*, whose factor supplies of capital and labor are ten times as large as those of a benchmark country, country *B*. Assuming that the two countries share the same Cobb–Douglas production function and have the same level of technology, this implies that, if the degree of returns to scale is 1.3, the ratio of labor productivity in country *A* to country *B* will be 2:1. If country *A*'s factor supplies are fifty times those of country *B* (e.g. the US vis-à-vis Singapore) then the predicted ratio increases to over 3:1.<sup>40</sup> The fact that Singapore has a level of productivity that is not very different from that of the United States suggests that even mild *aggregate* increasing returns are implausible. Alternatively, the scale problem implies that, given their similar levels of productivity, the level of technology in Singapore must be approximately three times that in the US.

How is the contrary argument, that constant returns are implausible as we do not observe what Krugman (1991a) terms “backyard capitalism,” where everyone practices self-sufficiency, answered? One answer lies in Samuelson's (1967) *asymptotic homogeneity theorem* which shows that if the increasing returns are the result of indivisibilities, then as output grows so the returns to scale will converge to constant returns to scale. At a small scale of production there may be substantial economies of scale preventing backyard capitalism, but these disappear as the scale of production increases. In this case, differences in firm, or even city, size can be explained purely by a stochastic or random process such as the law of proportionate effect (i.e. by Gibrat's law).

Kaldor (1966) despite his skepticism toward econometrics, but consistent with his inductive approach to theorizing, presented a major empirical challenge to the assumption of constant to scale. In his inaugural lecture on taking up a personal chair at the University of Cambridge in 1966 he addressed the question as to why the post-war economic performance of the UK had been so poor relative to most of the other advanced countries. Central to his explanation was the Verdoorn law.<sup>41</sup> This is an empirical relationship between the growth of industrial productivity and output (McCombie *et al.* 2002), which takes the form:

$$\hat{P} = a + b\hat{Q} \quad (1)$$

where *P* and *Q* are labor productivity and output respectively and a  $\hat{\phantom{x}}$  over a variable denotes a growth rate. If an explicit allowance for capital accumulation is made, the law is specified as:

$$T\hat{F}P = c + b'\hat{Q} \quad (2)$$

where  $T\hat{F}P = \hat{Q} - a\hat{K} - (1 - a)\hat{L}$  is the growth of total factor productivity and *a* is capital's share in total output. Recent studies of the Verdoorn law further extend



the specification both through the use of spatial econometric techniques which allow for the inclusion of neighbor spillover effects and through the inclusion of such additional variables as a measure of spatial production density and a proxy for a country's (region's) initial level of technology (Angeriz *et al.* 2008a, 2008b).

When equation (1) is estimated using cross-regional or cross-industry data,<sup>42</sup> the slope coefficient is found to be around one-half, which Kaldor interpreted as evidence of substantial returns to scale. (Equation (2) gives very similar results.) Although the estimation of dynamic Verdoorn law is subject to a number of specification and estimation issues (Angeriz *et al.* 2008a), the estimated Verdoorn coefficient ( $b$  in equation (1)) is generally found to be robust, implying the existence of substantial increasing returns, often in the order of 1.50.<sup>43</sup> Some of the more recent studies are reported in Table 2.2. These returns to scale are encompassing in that they include induced technical change and both dynamic and static returns to scale. But how are these results to be reconciled with the estimates of conventional production function studies, especially since, under certain assumptions, the Verdoorn law can be derived from the conventional Cobb–Douglas production function (Black 1962)? In fact, if the same dataset is used, the estimation of the static Verdoorn law in log-level form (or the Cobb–Douglas production function also in log-levels) either gives very small estimated increasing returns or constant returns to scale (see Table 2.2). This result is known as the Static–Dynamic Verdoorn paradox.

Table 2.2 Estimate of the Verdoorn law

<i>Study</i>	<i>Spatial coverage</i>	<i>Dynamic law</i>	<i>Static law</i>
Kaldor (1966)	OECD	IRS	n.a.
McCombie (1982)	OECD	IRS	CRS
Michl (1985)	OECD	IRS	n.a.
McCombie and de Ridder (1984)	US states	IRS	CRS
McCombie (1985)	US states <sup>a</sup>	IRS	n.a.
Bernat (1996)	US states	IRS	n.a.
Fingleton and McCombie (1998)	EU regions	IRS	CRS
Hansen and Zang (1996)	China, regions	IRS	n.a.
León-Ledesma (1999, 2000)	Spain, regions	IRS	CRS/IRS <sup>b</sup>
Angeriz <i>et al.</i> (2008a)	EU regions	IRS	CRS
Angeriz <i>et al.</i> (2008b)	EU regions <sup>c</sup>	IRS	CRS
		Time-series	
Harris and Lau (1998)	UK regions	IRS	
Harris and Lau (1999)	International data	IRS	

Notes

IRS denotes increasing returns to scale and CRS, constant returns to scale.

n.a. denotes not available.

a 17 industries.

b Depends upon the specification.

c 7 industries.



This paradox, which has important consequences for the estimation of the degree of returns to scale, can be explained by the existence of *spatial aggregation bias* in the estimation of the static law (McCombie and Roberts 2007). Regional data are often for spatial areas that have been delineated for administrative rather than economic reasons (e.g. the NUTS regions or the US states). Suppose the correct spatial unit of observation to determine the degree of returns to scale is what may be termed the *Functional Economic Area* (FEA), the extent of which is determined by, say, travel to work times. If there are several FEAs in each of the regions, as the FEA data on output and the factor inputs are summed arithmetically to get the regional (state) data, then it can be shown that using the log-levels will give estimates which are severely biased downwards. Thus, even if the underlying data-generating process is characterized by substantial increasing returns at the FEA level, these will not be detected. However, using aggregate growth rates, which are dimensionless numbers, largely avoids this problem and will give a more accurate estimate of the degree of returns to scale. Thus, the dynamic Verdoorn law is the correct specification and the evidence in favor of increasing returns is compelling (McCombie and Roberts 2007).

Table 2.3 shows a simple example that illustrates this. As *TFP* is defined as output divided by the weighted level of total factor inputs (*TFI* or  $K^a L^{(1-a)}$ ), i.e.  $TFP \equiv Q/K^a L^{(1-a)}$ , we may, from equation (2), write the “true” static Verdoorn law as  $TFI = \Lambda Q^{1-b'} = Q^{0.5}$  where  $\Lambda$  is a constant. (For ease of exposition, we assume no exogenous technical change.)

There are three regions: region *A* contains one FEA; region *B* has two; and region *C* has three. Each hypothetical FEA is of equal size and, at this level, the true value of  $b' = 0.5$ . Normalizing the data, we shall assume that, for each FEA,  $TFI = 10$ ,  $\Lambda = 1$  and  $Q = 100$ . The data for each FEA in regions B and C are summed, as they are in practice, arithmetically. There are two years,  $t = 1$  and  $t = 2$ , and *TFI* and output grow in each region at 2.5 percent and 5 percent between these years. If we were to use the growth rates for the three regions to estimate the Verdoorn law then we would uncover the true value of the Verdoorn coefficient, i.e.  $b' = 0.5$ .<sup>44</sup> However, if we were to pool the data for the two years and estimate the static Verdoorn law with a dummy variable to allow the intercept to shift, we would incorrectly estimate  $b' = 0$  which is not equal to the true value of the Verdoorn coefficient. The biased estimate of the static Verdoorn coefficient erroneously indicates constant returns to scale. Consequently, to recover an unbiased estimate of  $b'$  the Verdoorn law should be specified in terms of growth rates.

This explains why conventional estimates of the aggregate production function using regional or cross-industry data generally find constant returns to scale, while the use of growth rates in the dynamic Verdoorn law finds substantial increasing returns to scale.

Consequently, there seems to be convincing macroeconomic empirical evidence of the existence of increasing returns to scale, which substantiates the microeconomic evidence of, among others, Pratten (1971).

Table 2.3 An example of spatial aggregation bias

	TFI			Q			
	<i>t</i> = 1	<i>t</i> = 2	growth rate (%)	<i>t</i> = 1	<i>t</i> = 2	growth rate (%)	
Region A	10	10.25	2.5	100	105	5	1 FEA
Region B	20	20.50	2.5	200	210	5	2 FEAs
Region C	30	30.75	2.5	300	315	5	3 FEAs

Note

*t* denotes a time period. See text for definitions of TFI, Q and FEA.

## The new economics of increasing returns

Despite the historically ambivalent attitude of mainstream economics towards the concept of increasing returns, there has, as we have on occasion mentioned, been a radical transformation within several fields of economics over the past two decades. This transformation originated in the late 1970s in the field of industrial organization. In particular, in pioneering work, Dixit and Stiglitz (1977) provided the necessary tools to allow economists to model internal increasing returns in a manner considered acceptable to the mainstream. That is to say, in a manner that permits the maintenance of a tractable, but non-Walrasian, general equilibrium framework. Subsequently, the use of the Dixit-Stiglitz monopolistic competition framework was to be incorporated into trade theory beginning from the early 1980s with the work of Paul Krugman and the development of strategic trade theory (notable contributions include Krugman (1979, 1980); Ethier (1982); and Helpman and Krugman (1985)).

Following this, the revolution moved on to the field of growth theory, which had lain stagnant as a field of mainstream research since the mid-1960s (Barro and Sala-i-Martin 2004: 17). In particular, Paul Romer (1986) transformed growth theory, introducing increasing returns as a central component into mainstream explanations of the long-run growth process. Although, initially, Romer, following on from Arrow (1962) and Sheshinski (1967), was to model increasing returns as emanating solely as an externality, this was done more as an interim solution until he was able to deal satisfactorily with the market structure implications of assuming internal increasing returns by grafting the Dixit-Stiglitz framework on to growth theory (see Romer (1994: 11–19) on this point).

Finally, since the early 1980s, the Dixit-Stiglitz framework has contributed in no small measure to the emergence of economic geography, for the first time, as an area of serious research for mainstream economists. In particular, building on his strategic trade theory work, Krugman (1991a, 1991b) pioneered the “new economic geography” (NEG), also referred to by some as “geographical economics” (Brakman *et al.* 2009). This work, which sees increasing returns as *the* central element in the explanation of the spatial distribution of economic activity, has steadily gained in popularity and is also now in the process of being

mainstreamed into the thinking of both national and international policy-making agencies.<sup>45</sup>

Out of the aforementioned developments, it would seem to be the NEG which comes closest to exhibiting a lineage with the earlier work of, in particular, Myrdal (1957), Hirschman (1958) and Kaldor (see, *inter alios*, Kaldor 1970, 1972, 1985). Thus, whereas in endogenous, or “new,” growth theory, the analysis remains firmly rooted in the supply-side of the economy, this is not the case with the NEG. Rather, the NEG embraces the logic of “circular and cumulative causation” that results from interactions between the supply- and demand-sides of the economy, and, indeed, is replete with references to the phrase “cumulative causation.” Furthermore, the key figures in the NEG literature have explicitly acknowledged these earlier precedents in their own work (Krugman 1991a, 1995; Fujita *et al.* 1999), and the influential Krugman–Venables model in which spatial agglomeration results from linkages between upstream and downstream work has been presented as a formalization of Hirschman (see Brakman *et al.* (2009) on this point).

One natural question which therefore arises is that of what is the value-added of the NEG, apart from that of mainstreaming old, albeit very important, ideas? The main answer to this question is that, rather than assuming, even if based on the observation of reality, the existence of agglomeration economies, the NEG explicitly models the sources of such economies. Thus, instead of incorporating increasing returns in the form of a highly aggregative empirical relationship such as Verdoorn’s law, the core NEG model of Krugman (1991a, 1991b), for example, models such economies as the endogenous outcome of the interaction of the fixed costs of establishing a manufacturing plant with labor mobility and transport costs. As such, the agglomeration economies are explicitly linked to structural features of a model economy, implying that their strength is also endogenous to these features. This being the case, the effect of agglomeration economies may be expected to vary with an economy’s level of development, and, in particular, the level of transport costs.<sup>46</sup> Furthermore, in a typical NEG model, the agglomerating force of increasing returns is counteracted by one or more potential centrifugal forces. These centrifugal forces may take, for example, the existence of a dispersed and geographically immobile segment of the labor force (Krugman 1991a, 1991b) and/or congestion in the housing and other markets of the economy (see e.g. Thomas 1996; Hanson 2005). This systematic incorporation of dispersion forces alongside agglomeration forces in a coherent theoretical framework contrasts with Kaldor’s model of circular and cumulative causation, in which they appear only as an after-thought (Kaldor 1970). Indeed, in this regard, the NEG overcomes the criticism of the Kaldorian cumulative causation framework that it is characterized by “too much cumulation” (Gordon 1991).

In addition, but by virtue of its theoretical foundations, the NEG has lent itself to a body of empirical work which is building the evidence base for the widespread existence of various forms of agglomeration economies in both developed and developing countries alike (see e.g. the survey of empirical evidence on agglomeration economies by Rosenthal and Strange 2004).

Although it is difficult to deny the strengths of the NEG framework, it is, nevertheless, subject to a number of shortcomings from a Kaldorian methodological perspective. Thus, the NEG continues to be based on the axiomatic-deductive approach to modeling, and, as such, continues to make what for Kaldor are “unscientific” assumptions. These assumptions inevitably include those of utility and profit-maximizing behavior on the behalf of households and firms respectively. Indeed, it is these assumptions which help to qualify NEG as being neo-classical, and which, for the mainstream, are seen as a source of strength. Likewise, just as Kaldor criticized Walrasian general equilibrium theory for not showing how such axiomatic concepts as goods and processes of production “are to be defined or recognized in relation to empirical material” so it is the case with important features of the standard NEG framework (Kaldor 1972: 1238). For example, although, in the NEG framework, transport costs represent a central element in the endogenous generation of agglomeration economies, they adopt a form which is unrecognizable in reality. Specifically, NEG models lack any explicit transport sector. Rather, transport costs occur in the form of so-called “iceberg” transport costs, whereby a proportion of a good is taken to “evaporate” in shipment. Moreover, for reasons of analytical convenience, this proportion is taken *exponentially* to increase with the distance over which the good is shipped. However, such a functional form contradicts the known empirical facts regarding the nature of transport costs. In particular, it contradicts evidence relating to the existence of economies of both distance and scale in transport (see, *inter alios*, Bayliss and Edwards 1970; Jansson and Shneerson 1987; Laird *et al.* 2005: 539).<sup>47</sup> Although attempts are being made to incorporate more realistic functional forms for transport costs into the NEG, as, for example, with the work of Bröcker (2002) and Fingleton (2005), the resulting models are not analytically tractable and, as a consequence, numerical simulation techniques have to be relied upon to study their properties. From a Kaldorian perspective at least, this raises questions of whether the NEG is in danger of running afoul of the same pitfall that Kaldor identified Walrasian general equilibrium theory as having encountered: namely that of constructing “scaffolding [that] gets thicker and more impenetrable with every successive reformulation of the theory, with growing uncertainty as to whether there is a solid building underneath” (Kaldor 1972: 1239). Although, given that it is starting from a firmer foundation, this risk seems less pronounced for NEG theory than for Walrasian general equilibrium theory, Kaldor’s words are nevertheless ones which NEG theorists would do well to keep in mind as they advance in their modeling.<sup>48</sup>

In addition to the above, it is doubtful whether the NEG, while capturing the emphasis of Kaldor and others on the importance of increasing returns, also captures his conception, as well as that of Paul David, of the true nature of path-dependence. Hence, under appropriate configurations of parameters, NEG models are characterized by multiple equilibria. As a corollary, initial conditions can help to determine the trajectory and final outcome of an economy. In particular, in the simple core NEG model, if the fixed costs of establishing a manufacturing plant and the size of the manufacturing sector are large enough relative to

the level of transport costs, the initial split of manufacturing activity between two regions will determine which region “wins” in the sense of becoming fully agglomerated at the expense of the other region. A further implication of this is that sufficiently large shocks to a spatial economy can have hysteretic impacts on the spatial distribution of economic activity. This has led to empirical work which seeks to identify such shocks and look for hysteretic effects as a means of testing the NEG approach (see, most notably, Davis and Weinstein 2002). Notwithstanding the evident merits of such empirical work, it nevertheless remains the case that in NEG models “history matters” only insofar as initial conditions matter. However, to restrict history to the importance of initial conditions is, obviously, to take a rather restrictive view of its importance in determining spatial economic outcomes (Roberts 2002, 2007) and is clearly inconsistent with Kaldor’s emphasis on the importance of historical, as opposed to logical, time.

All of this brings us back to the debates between both Kaldor and Hahn, and David and Liebowitz and Margolis. In particular, because they are deterministic, NEG models are characterized by determinate equilibria. That is to say, provided that we enlarge Kaldor’s definition of “data” to include knowledge of the initial conditions, his statement that “the system will converge on a *unique point* the exact nature of which [...] can be deduced from the ‘data’” (Kaldor 1972: 393; emphasis in original) can be taken to apply equally to NEG theory as to Walrasian general equilibrium theory. Likewise, although “history matters” in NEG models, they clearly fail to exhibit path-dependence in David’s meaning of the phrase. This is not to necessarily decry the value of such models, but it does point to the need for plurality in both methodology and models if this particular aspect of spatial development processes is to be captured. In particular, it calls for a tolerance of methods which, while not necessarily meeting the defined standards of “rigor” of mainstream economics with regard to, for example, the modeling of market structure and the need to ensure model closure, nevertheless allow for potentially compelling insights through the modeling of spatial growth processes as non-ergodic. In this sense, extension of the more “old-fashioned” class of cumulative causation models, which are in the spirit of Dixon and Thirlwall (1975), provides potential for the capturing of such processes, as has been demonstrated, for example, by Setterfield (1997a, 1997b) and further explored by Roberts (2007).<sup>49</sup> Likewise, the location choice models of Arthur (1994), which have been largely forgotten by the spatial economics literature, may be worthy of renewed attention. Finally, it points to the importance of historical case studies in the understanding of contemporary urban and regional, not to mention national, economies to accompany formal econometric work.

## Notes

- 1 In 1820, GDP per capita in Africa was \$420, while in Western Europe it was \$1,243 (constant 1990 international dollars; Maddison 2001).
- 2 The example was inspired by an article in Diderot’s *Encyclopédie* which was first published in France in 1751.

- 3 The factory system superseded the putting-out system, whereby individual workers worked with their own equipment in their cottages. In contrast to the factory system, the possibilities for the division and supervision of labor were severely constrained under this system.
- 4 These theories, particularly the new economic geography, will be discussed in more detail in the final section of the chapter.
- 5 It could be argued that the priority should go to Von Thünen (1826).
- 6 As John Bates Clark (1899: 3; emphasis in the original) put it:

We may now advance the more general thesis – later to be proved – that, *where natural laws have their way, the share of income that attests to any productive function is gauged by the actual product of it*. In other words, free competition tends to give to labor what labor creates, to capitalists what capital creates, and to entrepreneurs what the coordinating function creates.

- 7  $\pi$  can be derived from the elasticity of demand for an industry's product in a neoclassical optimization model.
- 8 Senghaas (1985) has shown that the US and continental Europe used extensive protectionist measures in the early stages of industrialization in their attempt to compete with the UK's head start.
- 9 This Marshallian defence of perfect competition has, from time to time, been rediscovered (see e.g. Buchanan and Yoon 1999).
- 10 Recently, this idea has attracted renewed attention in the new economic geography.
- 11 Conlisk (1968) provided an early generalization to incorporate increasing returns in the production function, albeit still with diminishing returns to capital and labor individually. In this generalization, however, increasing returns do not, except under the most unlikely circumstances, affect the steady-state growth properties of the model provided the underlying technology is a Cobb–Douglas production function. Consistent with the general neglect of increasing returns, Conlisk's model attracted little attention.
- 12 Endogenous growth theory originates from the seminal paper of Romer (1986) and Robert Lucas's 1985 Marshall Lectures at the University of Cambridge, published as Lucas (1988).
- 13 Lamfalussy (1961) made a similar argument with his notion of "defensive investment" as opposed to "enterprise investment." The former is investment added to existing machinery as and when it wears out and often at the last moment, whereas the latter is coordinated investment in, for example, new plants. The rate of return on "enterprise investment" is greater than on ad hoc "defensive investment."
- 14 They are "chance events" in the sense that they cannot be predicted and occur randomly.
- 15 As evidence of this dissatisfaction Kaldor made reference to the 1971 Presidential addresses to the Royal Economic Society and section F of the *British Association* by E. H. Phelps Brown and G. D. N. Worswick respectively (Kaldor 1972: 1240; 1975: 399).
- 16 It should be emphasized that it was these neoclassical constructs that Kaldor was critical of, not the concept of equilibrium per se, which he used in, for example, his two-sector growth model (McCombie and Roberts 2008).
- 17 Kaldor (1972: 1242) cites Young (1928: 520): "It would be wasteful to make a hammer to drive a single nail; it would be better to use what ever awkward implement lies conveniently to hand."
- 18 This model was subsequently formalized by Dixon and Thirlwall (1975).
- 19 [Young 1928: 530].
- 20 [Young 1928: 528].
- 21 [Young 1928: 528].

- 22 The footnote numbers have been renumbered to make them consistent with those in this chapter.
- 23 Kaldor also argued that, even in focusing upon their allocative functions, Walrasian general equilibrium theory gets the story of how markets work wrong. In particular, he argued that in order to understand the reality of resource allocation one has to explicitly consider the role of “professional intermediaries who are both buyers and sellers simultaneously [and] without whom markets as an institution could hardly function” (Kaldor 1996: 10).
- 24 General equilibrium theory does not postulate linear–homogeneous and continuously differentiable production functions, nor does it necessarily assume perfect foresight as Kaldor asserted. However, as Hicks (1989: 12) points out, Walras  

does not have a production function, even a Ricardian production function; he just has a matrix of coefficients, which are stated to be independent of outputs so that they obey the rule of CRS [constant returns to scale]. [...] There is not much difference between [the] later form of Walras’s model and the all-round marginal productivity doctrine; it is all CRS.
- 25 Compare this with the claim made in the title of David’s (2007) paper: “Path Dependence: A Foundational Concept for Historical Social Science.”
- 26 As Arthur (1994: 25) pointed out, the Florence cathedral clock designed in the year 1443 moves counter-clockwise around a 24-hour clock face.
- 27 Arthur assumes that once made, adoption choices are irreversible.
- 28 This is the assumption made by Liebowitz and Margolis (1995: 215). It may be thought of as being due to some form of externality. An alternative assumption is that subsequent improvements to the payoffs do not accrue to the earlier adopters.
- 29 Assuming, for expositional purposes, there is no discounting.
- 30 It has been argued that VHS became the industry standard even though Betamax was technically superior (see Cusumano *et al.* 1992).
- 31 The early loss that this would entail is of no concern to the present cohort. As Jevons (1871) pointed out, “bygones are forever bygones.”
- 32 It would be path-dependent if, for example, the values of the income elasticities were affected by the past growth of the economy. For example, a slow growth rate may lead to the introduction of possibly draconian economic policies to improve the ratio of the income elasticities. Again, the emphasis is “may” – there is no necessity that this will come about. Whether the requisite policies are introduced will depend upon historical contingencies (Roberts 2007).
- 33 Increasing returns is not the only cause of path-dependence. Arthur (1988) also notes that, in addition to scale economies in production, path-dependence can be caused by network externalities and technological interrelatedness. It can also be caused by agglomeration economies.
- 34 A more complex stochastic process that Arthur (1994) uses in several different contexts, including the modeling of firm location choices, is the non-linear Polya process. With this process, the probability of any particular choice made is a function of the proportion of agents who have previously made that choice.
- 35 “Explained” would be a more accurate word than “predicted.”
- 36 That is to say, neoclassical economics generally makes a “certainty equivalence” assumption, although, it should be noted, more recent models in the field of monetary economics have relaxed this assumption (for a discussion of this see Roberts forthcoming).
- 37 The data Romer refers to is the relationship between both the growth in income per capita and the share of investment in GDP and the level of income per capita relative to the United States. However, he could just as easily have been referring to more general production relationships.



38 A similar functional form had, however, been used by Wicksell in his theoretical work.

39 See Douglas (1976) for a retrospective view of his studies.

40 These calculations follow from the formula

$$\frac{Y_i/L_i}{Y_0/L_0} = \frac{A(\lambda_i K_0)^{va} (\lambda_i L_0)^{(1-a)v-1}}{A(K_0)^{va} (L_0)^{(1-a)v-1}} = \lambda_i^{v-1}$$

where  $Y$ ,  $A$ ,  $K$  and  $L$  denote the levels of output, technology, capital and labor respectively;  $\lambda_i$  is the scale factor,  $a$  is capital's share and  $v$  is the degree of homogeneity of the common production function.

41 This relationship was postulated by P. J. Verdoorn (1949) in *L'Industria*, an Italian journal.

42 This is sometimes known as Fabricant's law. Kennedy (1971) finds strong evidence for increasing returns to scale using cross-industry data.

43 There are also likely to be returns to scale in the tertiary industries, but the fact that the measure of output is often calculated as the value of the inputs with an arbitrary allowance for productivity change means that any estimation results are likely to be misleading.

44 Of course, for meaningful estimation in practice we would require data for more than three regions. The restriction here to three regions is purely for expositional purposes. For a full and formal statement of the spatial aggregation bias argument, which allows for many regions, see McCombie and Roberts (2007).

45 See e.g. The World Bank (2008).

46 In particular, a characteristic prediction of one class of NEG models is that of the existence of an inverted U-shaped relationship between the general level of transport costs in an economy and the degree of inequality in the spatial distribution of economic activity (see e.g. the models of Fujita and Krugman (1995), and Krugman and Venables (1995)).

47 This assumption of exponential iceberg transport costs has been inherited from strategic trade theory. As such, similar criticisms may be made of this literature.

48 Similar to its treatment of transport costs, the NEG has been criticized for its use of the all-important Dixit–Stiglitz framework of monopolistic competition. These criticisms have taken place on the grounds that the type of competition implied leaves no room for strategic interaction between firms, when there are good reasons for believing that such interactions are central to the location decisions of firms (Neary 2001). Indeed, from a Kaldorian methodological perspective, fears of shaky foundations on this front would seem to be confirmed by Krugman's admission that the Dixit–Stiglitz framework is, in his own words, "completely unrealistic" (Krugman 1995: 60). However, it is worth noting that, subsequent to the criticism of Neary, alternative NEG models embodying alternative forms of competition, which do allow for strategic interaction, have been developed. See, for example, Combes and Lafourcade (2001), who instead assume Cournot competition.

49 See also Roberts and Setterfield (2007) on this point.

## References

- Angeriz, A., McCombie, J. S. L. and Roberts, M. (2008a) "Some New Estimates of Returns to Scale for EU Regional Manufacturing 1986–2002," *International Regional Science Review*, 31: 62–87.
- Angeriz, A., McCombie, J. S. L. and Roberts, M. (2008b) "Returns to Scale and Regional Growth. A Disaggregated Approach for Manufacturing in the EU Regions," Cambridge Centre for Economic and Public Policy, Department of Land Economy, University of Cambridge (mimeo).



- Arrow, K. J. (1962) "The Economic Implications of Learning by Doing," *Review of Economic Studies*, 29: 155–73.
- Arrow, K. J. (2000) "Increasing Returns: Historiographic Issues and Path Dependence," *European Journal of History of Economic Thought*, 7: 171–80.
- Arthur, W. B. (1988) "Competing Technologies: An Overview," in G. Dosi, C. Freeman, R. Nelson, G. Silverberg and L. Soete (eds) *Technical Change and Economic Theory*, London: Pinter.
- Arthur, W. B. (1989) "Competing Technologies, Increasing Returns, and Lock-in by Historical Events," *Economic Journal*, 99: 116–31.
- Arthur, W. B. (1994) *Increasing Returns and Path Dependence in the Economy*, Ann Arbor: University of Michigan Press.
- Barro, R. and Sala-i-Martin, X. (2004, 2nd edn) *Economic Growth*, Cambridge, MA: MIT Press.
- Bayliss, B. T. and Edwards, S. L. (1970), *Industrial Demand for Transport*, London: Ministry of Transport, HMSO.
- Bernat, G. A. (1996) "Does Manufacturing Matter? A Spatial Econometric View of Kaldor's Laws," *Journal of Regional Science*, 36: 463–77.
- Black, J. (1962) "The Technical Progress Function," *Economica*, 29: 166–70.
- Brakman, S., Garretson, H. and van Marrewijk, C. (2009) *The New Introduction to Geographical Economics*, Cambridge: Cambridge University Press.
- Bröcker, J. (2002) "Spatial Effects of European Transport Policy: A CGE Approach," in G. J. D. Hewings, M. Sonis and D. Boyce (eds) *Trade, Network and Hierarchies: Modeling Regional and Inter-regional Economies*, Advances in Spatial Science, Berlin: Springer.
- Buchanan, J. M. and Yoon, Y. J. (1999) "Generalized Increasing Returns, Euler's Theorem, and Competitive Equilibrium," *History of Political Economy*, 31: 511–23.
- Chamberlin, E. H. (1933) *The Theory of Monopolistic Competition*, Cambridge, MA: Harvard University Press.
- Clapham, J. H. (1922) "Of Empty Economic Boxes," *Economic Journal*, 32, 127: 305–14.
- Clark, J. B. (1899) *The Distribution of Wealth: A Theory of Wages, Interest and Profits*, New York: Macmillan.
- Cobb, C. W. and Douglas, P. H. (1928) "A Theory of Production," *American Economic Review*, 8, suppl.: 139–65.
- Combes, P.-P. and Lafourcade, M. (2001) "Transport Cost Decline and Regional Inequalities: Evidence from France," CEPR Discussion Paper, 2894.
- Conlisk, J. (1968) "Non-constant Returns to Scale in a Neoclassical Growth Model," *International Economic Review*, 9: 369–73.
- Cusumano, M. A., Mylonadis, Y. and Rosenbloom, R. S. (1992) "Strategic Manoeuvring and Mass-market Dynamics: The Triumph of VHS over Beta," *Business History Review*, 66: 51–94.
- David, P. A. (1975) *Technical Choice, Innovation and Economic Growth: Essays on American and British Experience in the Nineteenth Century*, Cambridge: Cambridge University Press.
- (1985) "Clio and the Economics of QWERTY," *American Economic Review* (Papers and Proceedings), 75: 332–7.
- (1986) "Understanding the Economics of QWERTY: The Necessity of History," in W. N. Parker (ed.) *Economic History and the Modern Economist*, Oxford: Oxford University Press.
- (1997) "Path Dependence and the Quest for Historical Economics: One More

- Chorus of the Ballad of QWERTY," University of Oxford Discussion Papers in Economic and Social History, 20.
- (1999) "At Last, a Remedy for Chronic QWERTY-Skepticism!" Working paper, All Souls College, University of Oxford.
- (2001) "Path Dependence, its Critics and the Quest for 'Historical Economics'," in P. Garrouste and S. Ioannides (eds) *Evolution and Path Dependence in Economic Ideas*, Cheltenham: Edward Elgar.
- (2007) "Path Dependence: A Foundational Concept for Historical Social Science," *Clometrica, Journal of Historical Economics and Econometric History*, 1: 91–114.
- Davidson, P. (1972) *Money and the Real World*, London: Macmillan.
- Davis, D. R. and Weinstein, D. E. (2002) "Bones, Bombs and Breakpoints: The Geography of Economic Activity," *American Economic Review*, 92: 1269–89.
- Dixit, A. K. and Stiglitz, J. E. (1977) "Monopolistic Competition and Optimum Product Diversity," *American Economic Review*, 67: 297–308.
- Dixon, R. and Thirlwall, A. P. (1975) "A Model of Regional Growth Rate Differences on Kaldorian Lines," *Oxford Economic Papers*, 27: 201–14.
- Dobb, M. (1973) *Theories of Value and Distribution since Adam Smith: Ideology and Economic Theory*, Cambridge: Cambridge University Press.
- Douglas, P. H. (1976) "The Cobb–Douglas Production Function Once Again: Its History, Its Testing, and Some New Empirical Values," *Journal of Political Economy*, 84: 903–15.
- Ethier, W. J. (1982) "National and International Returns to Scale in the Theory of International Trade," *American Economic Review*, 72: 389–405.
- Farrell, M. J. (1959) "The Convexity Assumption in the Theory of Competitive Markets," *Journal of Political Economy*, 57: 377–91.
- Fingleton, B. (2005) "Towards Applied Geographical Economics: Modelling Relative Wage Rates, Incomes and Prices for the Regions of Great Britain," *Applied Economics*, 37: 2417–28.
- Fingleton, B. and McCombie, J. S. L. (1998) "Increasing Returns and Economic Growth. Evidence for Manufacturing from the European Union Regions," *Oxford Economic Papers*, 50: 89–105.
- Fisher, F. M. (1992) *Aggregation. Aggregate Production Functions and Related Topics*, ed. J. Monz, London: Harvester Wheatsheaf.
- Fujita, M. and Krugman, P. R. (1995) "When is the Economy Monocentric?: von Thünen and Chamberlain Unified," *Regional Science and Urban Economics*, 25: 505–28.
- Fujita, M., Venables, A. J. and Krugman, P. R. (1999) *The Spatial Economy: Cities, Regions and International Trade*, Cambridge, MA: MIT Press.
- Gordon, D. (1991) "Kaldor's Macro System: Too Much Cumulation, Too Few Contradictions," in E. J. Nell and W. Semmler (eds) *Nicolas Kaldor and Mainstream Economics: Confrontation or Convergence?*, New York: St Martin's Press.
- Griliches, Z. and Mairesse, J. (1995) "Production Functions: The Search for Identification," National Bureau of Economic Research, Working Paper, 5067.
- Griliches, Z. and Ringstad, V. (1971) *Economies of Scale and the Form of the Production Function*, Amsterdam: North Holland.
- Hahn, F. (1973) *On the Notion of Equilibrium in Economics: An Inaugural Lecture*, Cambridge: Cambridge University Press.
- Hamilton, A. (1794) *Report on Manufactures*, reprinted in S. McKee (ed.) *Papers on Public Credit, Commerce and Finance*, New York: Columbia University Press, 1934.
- Hansen, J. D. and Zang, J. (1996) "A Kaldorian Approach to Regional Economic Growth in China," *Applied Economics*, 28: 679–85.

- Hanson, G. H. (2005) "Market Potential, Increasing Returns, and Geographic Concentration," *Journal of International Economics*, 67: 1–24.
- Harris, R. I. D. and Lau, E. (1998) "Verdoorn's Law and Increasing Returns to Scale in the UK Regions, 1968–91: Some New Estimates Based on the Cointegration Approach," *Oxford Economic Papers*, 50: 201–19.
- Harris, R. I. D. and Liu, A. (1999) "Verdoorn's Law and Increasing Returns to Scale," *Applied Economics Letters*, 6: 29–34.
- Helpman, E. and Krugman, P. R. (1985) *Market Structure and Foreign Trade: Increasing Returns, Imperfect Competition, and the International Economy*, Cambridge, MA: MIT Press.
- Hicks, J. R. (1989) "The Assumption of Constant Returns to Scale," *Cambridge Journal of Economics*, 13: 9–17.
- Hildebrand, G. and Liu, T. (1957) *Manufacturing Production Functions in the United States*, Ithaca, NY: Cornell University Press.
- Hirschman, A. O. (1958) *The Strategy of Economic Development*, New Haven, CT: Yale University Press.
- Jansson, J. O. and Shneerson, D. (1987) *Linear Shipping Economics*, New York: Chapman and Hall.
- Jevons, W. S. (1871) *The Theory of Political Economy*, London: Macmillan.
- Jones, C. I. (2001, 2nd edn) *Introduction to Economic Growth*, New York: Norton.
- Kaldor, N. (1966) *The Causes of the Slow Economic Growth of the United Kingdom: An Inaugural Lecture*, Cambridge: Cambridge University Press.
- (1970) "The Case for Regional Policies," *Scottish Journal of Political Economy*, 17: 337–48.
- (1972) "The Irrelevance of Equilibrium Economics," *Economic Journal*, 82: 1237–55.
- (1975) "What is Wrong with Economic Theory," *Quarterly Journal of Economics*, 89: 347–57.
- (1977) "Capitalism and Industrial Development: Some Lessons from Britain's Experience," *Cambridge Journal of Economics*, 1: 193–204.
- (1979) "Equilibrium Theory and Growth Theory," in M. J. Boskin (ed.) *Economics and Human Welfare, Essays in Honour of Tibor Scitovsky*, New York: Academic Press.
- (1985) *Economics without Equilibrium*, New York: M. E. Sharpe.
- (1996) *Causes of Growth and Stagnation in the World Economy* (Raffaele Mattioli Lectures, Milan, 1984), Cambridge: Cambridge University Press.
- Kennedy, K. A. (1971) *Productivity and Industrial Growth: The Irish Experience*, Oxford: Clarendon Press.
- Kirman, A. (1989) "The Intrinsic Limits of Modern Economic Theory: The Emperor has no Clothes," *Economic Journal*, 99: 126–39.
- Klotz, B. P. (1970) "Productivity Analysis in Manufacturing Plants," Bureau of Labor Statistics Staff Paper, 9, U.S. Department of Labor.
- Krugman, P. R. (1979) Increasing Returns, Monopolistic Competition, and International Trade, *Journal of International Economics*, 9: 469–79.
- (1980) "Scale Economies, Product Differentiation, and the Pattern of Trade," *American Economic Review*, 70: 950–9.
- (1991a) *Geography and Trade*, Cambridge, MA: MIT Press.
- (1991b) "Increasing Returns and Economic Geography," *Journal of Political Economy*, 99: 483–99.
- (1995) *Development, Geography and Economic Theory*, Cambridge, MA: MIT Press.

- Krugman, P. R. and Venables, A. J. (1995) "Globalisation and the Equality of Nations," *Quarterly Journal of Economics*, 110: 857–80.
- Laird, J. J., Nellthorpe, J. and Mackie, P. J. (2005) "Network Effects and Total Economic Impact in Transport Appraisal," *Transport Policy*, 12: 537–44.
- Lamfalussy, A. (1961) *Investment and Growth in Mature Economies: The Case of Belgium*, London: Macmillan.
- Lee, F. S. (1998) *Post-Keynesian Price Theory*, Cambridge: Cambridge University Press.
- León-Ledesma, M. A. (1999) "Verdoorn's Law and Increasing Returns: An Empirical Analysis of the Spanish Regions," *Applied Economics Letters*, 6: 373–6.
- (2000) "Economic Growth and Verdoorn's Law in the Spanish Regions, 1962–91," *International Review of Applied Economics*, 14: 55–69.
- Liebowitz, S. J. and Margolis, S. E. (1995) "Path Dependency, Lock-in, and History," *Journal of Law, Economics and Organization*, 11: 204–26.
- List, F. (1841, reprinted 1996) *National System of Political Economy*, New York: Augustus M. Kelley.
- Lucas, R. E. (1988) "On the Mechanics of Economic Development," *Journal of Monetary Economics*, 22: 3–42.
- McCombie, J. S. L. (1982) "Economic Growth, Kaldor's Laws and the Static–Dynamic Verdoorn Law Paradox," *Applied Economics*, 14: 279–94.
- (1985) "Increasing Returns and the Manufacturing Industries: Some Empirical Issues," *The Manchester School of Economic and Social Studies*, 53: 55–75.
- McCombie, J. S. L. and de Ridder, J. R. (1984) "The Verdoorn Law Controversy: Some New Empirical Evidence Using US State Data," *Oxford Economic Papers*, 36: 268–84.
- McCombie, J. S. L. and Roberts, M. (2007) "Returns to Scale and Regional Growth: The Static–Dynamic Verdoorn Law Paradox Revisited," *Journal of Regional Science*, 47: 179–208.
- (2008) "Effective-demand-constrained Growth in a Two-sector Kaldorian Model," *Journal of Post Keynesian Economics*, 31: 57–78.
- McCombie, J. S. L., Pugno, M. and Soro, B. (2002) *Productivity Growth and Economic Performance. Essays on Verdoorn's Law*, Basingstoke: Palgrave.
- Maddison, A. (2001) *The World Economy: A Millennial Perspective*, Paris: OECD Development Centre Studies.
- Mankiw, N. G., Romer, D. and Weil, D. (1992) "A Contribution to the Empirics of Economic Growth," *Quarterly Journal of Economics*, 107: 407–37.
- Marshall, A. (1890) *Principles of Economics*, 9th variorum, ed. C. W. Guillebaud, London: Macmillan, 1961.
- Mendershausen, H. (1938) "On the Significance of Professor Douglas' Production Function," *Econometrica*, 6: 143–53.
- Michl, T. R. (1985) "International Comparisons of Productivity Growth: Verdoorn's Law Revisited," *Journal of Post Keynesian Economics*, 7: 474–92.
- Moroney, J. R. (1972) *The Structure of Production in American Manufacturing*, Chapel Hill: University of North Carolina Press.
- Myrdal, G. (1957) *Economic Theory and Underdeveloped Regions*, London: Duckworth.
- Neary, J. P. (2001) "Of Hype and Hyperbolas: Introducing the New Economic Geography," *Journal of Economic Literature*, 39: 536–61.
- Pons-Novell, J. and Viladecans-Marsal, E. (1999) "Kaldor's Laws and Spatial Dependence: Evidence for the European Regions," *Regional Studies*, 33: 443–51.
- Pratten, C. (1971) *Economies of Scale in Manufacturing Industry*, Cambridge: Cambridge University Press.

- Ricardo, D. (1817) *On the Principles of Political Economy and Taxation*, Vol. 1 of *The Works and Correspondence of David Ricardo*, ed. P. Sraffa, Cambridge: Cambridge University Press, 1951.
- Roberts, M. (2002) *Essays in Cumulative Causation*, unpublished PhD thesis, Department of Land Economy, University of Cambridge.
- (2007) “Modelling Historical Growth: A Contribution to the Debate,” in P. Arestis, J. S. L. McCombie and R. Vickerman (eds) *Growth and Development. Essays in Honour of A.P. Thirlwall*, Cheltenham: Edward Elgar.
- (forthcoming) “The Theory and Practice of Monetary Policy,” in N. F. B. Allington and J. S. L. McCombie (eds) *Cambridge Handbook of Applied Economics*, Vol. 2, Cambridge: Cambridge University Press.
- Roberts, M. and Setterfield, M. (2007) “What is Endogenous Growth?,” in P. Arestis, M. Baddeley and J. S. L. McCombie (eds) *Understanding Economic Growth: New Directions in Theory and Policy*, Cheltenham: Edward Elgar.
- Robinson, J. V. (1933) *The Economics of Imperfect Competition*, London: Macmillan.
- Romer, P. M. (1986) “Increasing Returns and Long-run Growth,” *Journal of Political Economy*, 94: 1002–37.
- (1994) “The Origins of Endogenous Growth,” *Journal of Economic Perspectives*, 8: 3–22.
- Rosenstein-Rodan, P. N. (1943) “Problems of Industrialisation in Eastern and South Eastern Europe,” *Economic Journal*, 53: 202–11.
- Rosenthal, S. S. and Strange, W. C. (2004) “Evidence on the Nature and Sources of Agglomeration Economies,” in J. V. Henderson, and J. F. Thisse (eds) *Handbook of Regional and Urban Economics*, Amsterdam: Elsevier B.V.
- Samuelson, P. A. (1967) “The Monopolistic Competition Revolution,” in R. E. Kuenne (ed.) *Monopolistic Competition Theory: Studies in Impact*, New York: John Wiley & Sons.
- Senghaas, D. (1985) *The European Experience: A Historical Critique of Development Theory*, Leamington Spa: Berg Publishers.
- Setterfield, M. (1997a) “‘History versus Equilibrium’ and the Theory of Economic Growth,” *Cambridge Journal of Economics*, 21: 365–78.
- (1997b) *Rapid Growth and Relative Decline: Modelling Macroeconomic Dynamics with Hysteresis*, London: Macmillan.
- Sheshinski, E. (1967) “A Model of Inventive Activity and Capital Accumulation,” in K. Shell (ed.) *Essays on the Theory of Optimal Economic Growth*, Cambridge, MA: MIT Press.
- Smith, A. (1776, 1973) *The Wealth of Nations*, ed. A. Skinner, Harmondsworth: Penguin Books.
- Solow, R. M. (1956) “A Contribution to the Theory of Economic Growth,” *Quarterly Journal of Economics*, 70: 65–94.
- (1957) “Technical Change and the Aggregate Production Function,” *Review of Economics and Statistics*, 39: 312–20.
- Sraffa, P. (1926) “The Laws of Returns under Competitive Conditions,” *Economic Journal*, 36: 535–50.
- Starr, R. M. (1969) “Quasi-equilibria in Markets with Non-convex Preferences,” *Econometrica*, 37: 25–38.
- Stigler, G. (1951) “The Division of Labor is Limited by the Extent of the Market,” *Journal of Political Economy*, 59: 185–93.
- Swan, T. W. (1956) “Economic Growth and Capital Accumulation,” *Economic Record*, 32: 334–61.

- Thirlwall, A. P. and Dixon, R. (1979) "A Model of Export Led Growth with a Balance of Payments Constraint," in J. K. Bowers (ed.) *Inflation, Development and Integration*, Leeds: Leeds University Press.
- Thomas, A. (1996) "Increasing Returns, Congestion Costs and the Geographic Concentration of Firms," mimeo, Washington, DC: The International Monetary Fund.
- Toner, P. (1999) *Main Currents in Cumulative Causation. The Dynamics of Growth and Development*, Basingstoke: Macmillan.
- Veblen, T. (1915) *Imperial Germany and the Industrial Revolution*, New York: Macmillan.
- Verdoorn, P. J. (1949) "Fattori che Regolano lo Sviluppo della Produttività del Lavoro," *L'Industria*, 1: 3–10 (English translation by Anthony Thirlwall – Chapter 3 in *Productivity Growth and Economic Performance. Essays on Verdoorn's Law*, ed. J. S. L. McCombie, M. Pugno and B. Soro, Basingstoke: Palgrave.
- World Bank (2008) *World Development Report 2009, Reshaping Economic Geography*, Washington, DC: The World Bank.
- Young, A. (1928) "Increasing Returns and Economic Progress," *Economic Journal*, 38: 527–42.
- von Thünen, J. H. (1826, English translation 1966) *The Isolated State*, ed. P. Hall, English edition of *Der Isolierte Staat*, trans. C. M. Wartenberg, Oxford and New York: Pergamon Press.

### 3 Cumulative causation and Northeast Asian post-war industry policy

*Phillip Toner and Gavan Butler*

#### Introduction

The purpose of this chapter is to describe how the theory of circular and cumulative causation accounts for a number of key, commonly identified elements of Northeast Asian development strategy since 1950, especially the use of specific industry policies. Several leading figures in the cumulative causation (CC) tradition have suggested that the pattern of Northeast Asian development is consistent with the CC understanding of the growth process in industrial economies and, consequently, may provide empirical support to this theory (Kaldor 1966: 20; Cornwall 1977: 191–3; Eatwell 1982: 136–9; 1987: 737–8; McCombie and Thirlwall 1994). However, these leading figures have provided only brief suggestions on these matters and have neither undertaken detailed studies from a CC perspective of national development strategies generally nor of Northeast Asian industry policies in particular. This is explained in part by the dominance of econometric techniques as the principal methodology of CC theorists.<sup>1</sup> They have generally eschewed recourse to the empirical work of economic historians and political scientists on Northeast Asian development.

At the outset we acknowledge that the preponderant references are to Japan, the first Northeast Asian “developmental state.” It is widely understood, however, that the strategies of South Korea and Taiwan followed those of Japan, as is made clear in the studies cited in Part 2.

The central argument of this chapter is that there is a very large body of evidence, mainly from economic historians, political economists and political scientists who have studied the institutions and policies underpinning Northeast Asian development, which is strongly consistent with CC theory.

The studies included in this survey (in Section 4 of Part 2) may be broadly identified as supporting the “developmental state” or “governed market” interpretation of post-war Northeast Asian industrial development (see e.g. Johnson 1982; Wade 1990; Woo-Cumings 1999; Jomo 2001). This view ascribes a key role to the state and its economic agencies in planning and directing the industrial structure. There was also, however, a general consensus between industry and government over the means and ends of industry policy, with



considerable consultation and negotiation between the bureaucracy and businesses regarding policy formation and implementation. An important theme common to these writers is the inadequacy of neoclassical economic theory to comprehend and explain the goals and effects of post-war Northeast Asian industry policy, except that contradictorily they frequently allude to “market failure.” Generally, the studies draw on diverse heterodox sources to explain the reasons for, and effects of, Northeast Asian industry policies without establishing a satisfactory understanding of the economic rationale and effects of such policies. While these “statist” writers provide a detailed account of the institutions, goals, instruments and effects of industry policy they generally lack a coherent overarching theory to explain the phenomenon they describe. We argue that CC theory can provide this coherent explanation.

The chapter is divided into two parts. The first is an outline of fundamental CC principles. The second is an account of how these principles explain certain major elements (or “stylised facts” to use Kaldor’s term) of post-war Northeast Asian development in the second half of the twentieth century.

## **Part 1**

### ***The theory of circular and cumulative causation***

Perhaps the simplest description of CC theory is to say that it inverts all the major assumptions of neoclassical economics. It is concerned with the explanation of growth and dynamics within an industrial system, not the optimal allocation of fixed resources; indivisibilities and increasing returns are pervasive and very significant in contrast to the neoclassical assumption of constant returns in production and diminishing returns to a factor; technological change is endogenous, not exogenous to the system; complementarities in investment are far more important than the principle of substitution; there is no monotonic inverse relation between the rate of interest and capital–labor ratio, growth in the size of the market determines factor proportions with a larger volume of production associated with a higher capital–labor ratio; factors of production are not exogenously given but are “created” endogenously in the process of production; external economies are pervasive and significant; and finally, consumer preferences are not exogenous but alter systematically with the level of per capita income.

These opposite assumptions lead in turn to major differences in the understanding of the growth process and economic policy. Neoclassical economics emphasizes the equilibrating effects of the market through price and output changes and factor mobility and that economic growth is best secured by allocating resources to reflect current prices.

The following presents a fuller account of what arguably are the key elements in CC growth theory and the strategy for development advanced by the writers mentioned above.



*Increasing returns*

According to CC theory, growth in per capita output is due primarily to increasing returns. Increasing returns occur in manufacturing industries with primary and tertiary industries subject to diminishing and constant returns respectively (Kaldor 1966). CC theory assumes that increasing returns are due primarily to the division of labor or the specialization of production across industries and within firms. A crucial aspect of the division of labor is the employment of more capital-intensive production methods and more specialized or dedicated capital equipment. The increase in the capital–labor ratio is due to the overcoming of indivisibilities in the employment of capital goods and cheapening in the real price of capital goods, since as produced means of production these are also subject to cost reduction through increasing returns in their production. The major innovation of CC theory was to emphasize that increasing returns operate at an industry, inter-industry and economy-wide level and are a dynamic process: cumulative increases in the size of the market give rise to cumulative extension of the division of labor and increase in the capital–labor ratio. In Kaldor's (1966: 9) words, "increasing returns is a 'macro-phenomenon'."

Other sources of increasing returns are also important. These include static plant-based scale economies such as economies of massed resources, the law of physical dimensions and economies of buying and selling as a plant increases its output. Static plant-based economies confer a benefit to a firm which can be lost if the level of output of the firm declines.

Dynamic economies are also important and include "learning by using" and "learning by doing" (Kaldor 1972). Dynamic economies result from the accumulation of experience with production over time, in contrast to static plant-based scale economies which result from a once-and-for-all increase in the size of a plant or business. Not only does learning by doing arise from repetition of production tasks and result in productivity gains from increased familiarity with and dexterity in existing production systems, but such learning also leads to incremental modifications and improvements to these processes. Learning by using, also known as "user–producer interaction," entails feedback from users of intermediate, capital or final consumption goods to producers resulting in design and other modifications that extend the range of uses, improve performance or reduce the cost of the goods (Rosenberg 1982: 121–2; von Hippel 2005).

In line with the importance it attaches to dynamic economies, CC theory assumes that technological change is largely endogenous to the system, and is for a number of reasons (such as overcoming indivisibilities and high barriers to entry associated with product and process R&D) facilitated by growth in the size of the market (Young 1928: 535). CC theory rejects the neoclassical view of a clear-cut distinction between the effects of technical progress as represented by a shift in the production function and a movement along the function due to an increase in capital per worker (Kaldor 1972). Changes in factor proportions are one of the key

transmission mechanisms for technological change since more recent “vintages” of capital equipment embody more efficient production methods.

It is evident from the above that the CC view of increasing returns is catholic, embracing static and dynamic economies, endogenous technological change and changes in factor prices. This catholic view of increasing returns differs markedly from the “rigorous” neoclassical account of scale economies which is restricted to an analysis of changes in output under static conditions of equi-proportional changes in factor inputs (Gold 1981: 14).

To enable growth arising from increasing returns to be sustained, CC theory assumes that factor supply is elastic or, more specifically, that the rate of growth of factor supply is governed by the rate of growth of demand (McCombie and Thirlwall 1994). Capital goods are produced means of production and as such are themselves subject to increasing returns. Savings are not assumed to constrain investment as investment generates additional savings through the income multiplier and the propensity to save increases along with income levels. Labor flows from low to higher productivity industries and regions.

In summary, the growth of productivity and the growth of total output are linked in a dynamic, circular and cumulative relation. Growth in total output extends the scope for increasing returns in production, and the resulting increase in productivity and reduction in the real price of intermediate inputs, capital and consumer goods has the effect of widening the market. CC theory assumes that goods subject to increasing returns – that is, manufactures – have high price and income elasticities of demand. Thus a reduction in unit price will lead to a more than proportionate increase in demand; and, empirically, it has been shown that over a wide range of income levels the share of income spent on manufactures increases more than proportionately with increases in per capita income (Pasinetti 1981: 70). A simple model of growth is generated when it is assumed that manufactures are subject to both elastic supply (through increasing returns and endogenous factor creation) and elastic demand.

### *External economies*

External economies are benefits generated by some producers and consumers but rendered free to other producers and consumers. “Technological” externalities arise when there is “direct interdependence among members of the economy” such that the “output of the individual producer may depend upon the activities of other firms” (Scitovsky 1954: 288–99). These economies are not transmitted through the market mechanism via the exchange of goods and services. Examples of technological externalities include:

- economies of agglomeration, like the deepening of a skilled labor market and the freer flow of information about market opportunities (Meade 1952);
- low-cost access by firms to proprietary product and process innovations due to imperfect intellectual property rights governing these innovations (Romer 1994);

- development of national and international measurement, product and quality standards lowering costs and expanding the market for individual producers;
- investment by employers in training of workers in one industry who subsequently transfer to other industries, taking their skills with them.

There is another class of externality, pecuniary externalities, in which the profits of a firm may depend not only on its own output and investment decisions but also on the output and investment decisions of other firms or industries (Scitovsky 1954: 300). An example of pecuniary externality is the benefit afforded firm B when firm C also becomes a customer of firm A, enabling firm A to increase its output, exploit scale economies and reduce its output prices (Hirschman 1958). Network effects such as the benefit afforded to firms involved in e-commerce when other firms invest in Internet access are a more recent example.

### *Complementarities in production and investment*

For Kaldor, given his concern with growth and dynamics as opposed to the allocation of given resources, complementarities in production and investment are far more pervasive and significant than the neoclassical principle of substitution. Producer and consumer optimization with given resources requires ready substitution between factors and between consumption goods, respectively. The assumption of ready substitution is essential for convexity conditions and determinate, mathematically tractable solutions to “the economic problem.”

Complementarity in production is a function of indivisible factors and implies fixed factor coefficients in production. CC theory rejects the assumption of diminishing returns to a factor, especially capital, as increments to the capital stock are complementary with this stock. This complementarity arises from technological change which is embodied in newer vintages of capital goods. The technological change is a product of learning and, as emphasized by Allyn Young (1928), of the overcoming of indivisibilities in the development and deployment of more heterogeneous and specialized capital inputs. Finally, another important property of the heterogeneity of capital goods, especially different types of infrastructure, is that each type must be supplied jointly for the productivity potential of each to be realized.

## **The CC strategy for development**

The focus of CC theory is on growth and dynamic efficiency gains arising from increasing returns, endogenous technological change, complementarities in production and investment, and external economies.

### *Role of the state in coordinating investment decisions*

In the absence of a competitive equilibrium where market prices do not necessarily reflect present and future demand and supply conditions it can no

longer be assumed that private investment and consumption decisions are compatible with a socially optimal allocation of resources. Take the case of a large capital-intensive investment subject to increasing returns, such as those arising from static scale economies and learning by doing: such a plant has the potential to significantly lower costs to all user firms if it can expand. Yet there is no way the market can indicate that new customers would arise if costs were so lowered. The risk that no new customers may arise may well dissuade the supplier firm from expanding. The result is described by neoclassicists as a divergence between private and social costs and benefits. One method of reducing the risk faced by individual investors is for the state to coordinate investment decisions and demand conditions to ensure that the volume, timing and industrial composition of investments are linked in such a way as to ensure that the output of one plant can find a sufficient level of demand as a capital or intermediate input into another plant or into final consumption to validate the investment (Rosenstein-Rodan 1943: 251). Demand can be directed through measures such as tariffs, export subsidies and government procurement. By identifying the interdependence of investment decisions and by planning and sequencing the investments (using methods such as input-output analysis and linear programming) the state can ensure that risk is reduced as a barrier to investment and that increasing returns are exploited (Hirschman 1958; Chenery 1959). CC theory regards the reduction in risk following the coordination of investment decisions as a particularly important form of external economy (Rosenstein-Rodan 1943: 249–50).

### *Role of manufacturing industry in development*

Several criteria are used within CC theory and its programme for development to identify the dynamic growth potential of particular industries. These include:

- the income elasticity of demand for an industry's output;
- the degree to which plants, firms and industries are subject to increasing returns;
- the scope for technological improvement in an industry and the extent to which these technological improvements are embodied in the industry's outputs, notably as intermediate or capital inputs to other industries which raise the productivity or improve the quality of output of the purchasing industry;
- the complementarity between industries or relative magnitude of backward and forward linkages within and between industries.

CC theorists argue that manufacturing industry exhibits these criteria for dynamic growth potential, in contrast with agricultural and mining products. In addition, in contrast with primary activities, manufacturing largely generates its own inputs to production (be they machine tools or skilled labor) so that in the long run manufacturing output is not significantly supply constrained. Indeed,

given the high supply elasticity of manufacturing output, Kaldor (1972) argued that manufacturing output is “demand constrained” not “supply constrained.” Significant productivity growth rates, high supply elasticity in manufacturing output and high income elasticity of demand for manufactures create a virtuous circle of growth whereby a faster rate of growth of output generates faster rates of productivity and demand. Kaldor illustrated some of these relations in a series of statistical findings, which A. P. Thirlwall subsequently termed “Kaldor’s Growth Laws” (Kaldor 1966; Thirlwall 1983). The most famous of these Growth Laws states that a 1 per cent increase in manufacturing output is causally associated with a 0.5 per cent increase in manufacturing productivity. This is also termed the Verdoorn law, after its original proponent P. J. Verdoorn (McCombie *et al.* 2002).

Conversely, CC writers have emphasized the problems of dependence on primary commodities. Myrdal (1958: 52) observed that historically, demand for primary commodities was inelastic with respect to price, and moreover that primary commodities were subject to wide price fluctuations and adverse terms of trade.

### *Compatibility of import substitution and export-led growth*

For the above writers the focus of initial development effort is the growth of the domestic market. The input–output relations of poor nations, unlike rich industrial nations, are limited; that is, there is a small number of domestic intermediate and final demand industries and the intra- and inter-industry transactions are very restricted (Leontief 1963: 169–70). An underdeveloped economy is one that has a very large leakage of demand into imports. The structure of underdeveloped economies is such that the circular and cumulative mechanism of simultaneous expansion in demand and supply capacity is insufficiently developed for growth to occur. The goal of development policy is to expand and deepen a country’s input–output structure to create an interdependent or integrated industrial structure. This is more than a simple descriptive tautology. Rather, “filling-in” of the national transactions table creates a virtuous circle of growth (Hirschman 1958).

It must be emphasized that these same writers did not advocate autarky but advanced the proposition that the optimum strategy for increasing per capita income in underdeveloped economies was the growth of domestic output, initially through import substitution which, in turn, would be the basis of subsequent export-led growth. There were various reasons for concentrating the development effort on import substitution and the domestic market. They rejected comparative advantage as a basis for trade and development policy. Import substitution was argued to increase the size of the market, overcome indivisibilities in investment, extend the division of labor and accelerate learning by doing and using. Growth in the size of the market would permit cumulative increases in productivity creating a foundation for export competitiveness which in turn would increase the size of the market. Subsequently, Kaldor (1966, 1967;

Argyrous 1996) formalized this in a typology of development in his four-stage model of industrial development.

For a developing nation, exports are necessary to fund imports of capital goods and raw materials. Sufficient exports are necessary to avoid deflationary policies associated with persistent balance of payments deficits. Persistent balance of payments deficits impose a ceiling on domestic rates of growth. This “balance of payments constraint to growth” is a function of differences between a nation’s propensities to import and to export, which in turn reflect differences in the industrial structure of nations, especially the size and dynamism of their manufacturing industries (Kaldor 1978; Thirlwall 1979; McCombie and Thirlwall 1994). Exports also increase the “size of the market” available to domestic firms and therefore are an important source of increasing returns to the national economy.

This approach to development strategy founded in the post-war years of the twentieth century is supported by contemporary critical analyses of trade and growth theory. Deraniyagala and Fine (2001: 821), for example, in their critical review of the evidence on the benefits of “trade openness” have called for a rejection of the “crude dichotomies ... between free trade and protection” and argue for benefits of “sophisticated, sector- and country-specific trade and industry policy.” Nelson and Pack (1999) argued for the centrality of “learning” in industrial development and the role of the state in establishing the incentives to accelerate such learning. Because much technological knowledge is tacit and not codified it can only be acquired and refined through the production and development process. “[L]earning to operate effectively in the world of modern practice takes time and effort, the policy environment needs to nurture learning” (Nelson and Pack 1999: 435). They describe how governments in Taiwan and Korea provided technology diffusion services and protection from imports in return for the achievement of rates of productivity increase and product innovation that would open export markets. Access to these markets would in turn boost output and simultaneously the scope for industrial learning.

## **Part 2**

Three broad interpretations may be discerned in the literature on post-war Northeast Asian development. The neoclassical view attributes development to a fast rate of growth of factors of production and the conformity of production and trade with comparative advantage (Komiya and Itoh 1988; Saxonhouse 1988; World Bank 1991; Young 1995). This comparative advantage is not fixed but is “dynamic,” changing over time at unexplained rates with shifts in factor proportions and technology. Given the flexibility of prices, market forces prompt the allocation of factors to the most profitable and thus nationally desirable industries. Just what industries private investors choose to invest in does not matter, since prices reflect future private returns that are based on well-informed assessments of interrelationships between industries. From this perspective, state policies promoting selected industries are either ineffective or perverse in their impact on development (Argy and Stein 1997: 98–100). The state does, however, have a role

in macro-management by encouraging a high savings rate, low inflation, a sound education system and adequate infrastructure (Japan Economic Institute 1991).

The second interpretation emphasizes particular social values, culture and/or religion as being especially propitious to rapid economic development. The post-war performance of Japan, for example, is explained in terms of its particular national character – *nihonjinron* – such as respect for authority, strong nationalism and group orientation which, in turn, are embodied in practices such as lifetime employment, company unions and industrial organization by way of *keiretsu*, and consumer resistance to foreign goods.

The third interpretation, and the one that is the focus of this section, is that broadly associated with the notion of the “developmental state” (Johnson 1982; Wade 1990; Wade *et al.* 2004; Tyson and Zysman 1989; Amsden 1989, 2001; Yamamura 1990; Woo-Cumings 1991, 1999; Jomo 2005; Chang 2006). In these writers’ work, concepts of increasing returns, externalities and complementarities are either explicitly or implicitly identified in explaining the growth of Northeast Asian industrial economies. Moreover, a crucial role is identified for state industry policies in redressing market coordination failures in shaping and directing the pattern and pace of economic development. It is the central contention of this chapter that the explanation of the unprecedented post-war Northeast Asian growth provided by these developmental state writers is strongly consistent with CC theory. However, these writers generally provide no clear or coherent theoretical statement as to why or how the rate of aggregate growth reflects the selection of some industries for promotion over others. No overarching or coherent theoretical structure is offered to explain why the selection of industries matters.

The purpose of this section is to highlight three elements of Northeast Asian industry policy that are emphasized in this literature. It is argued that these three elements are consistent with CC theory and development strategy. The first element is so obvious it is in danger of being overlooked. Industry policy focused on the growth of manufacturing industry. The Northeast Asian countries explicitly rejected notions of static comparative advantage, which given their post-war factor proportions would have meant economies based on agriculture and simple labor-intensive handicrafts. The second key element is the crucial role of import substitution and of the growth of domestic demand in development. The third and final element is the promotion of state-selected industries to create an integrated industrial structure that would maximize backward and forward linkages in a virtuous circle of increasing output, incomes, investment, productivity and demand. The success of the instruments of industry policy can be explained through the selection of industries in accordance with the interdependence of their input–output and investment relations, exploiting increasing returns and “harvesting” and further creating a variety of external economies.

### ***First element: focus on manufacturing***

Various authors have highlighted the creation in Japan after the Second World War of a “consensus on the basic aims of industry policy” within the key



economic institutions, and among political parties and the populace (Eads and Yamamura 1988: 458; see also Stein 1993; Yamamura 1990; Johnson 1982; Johnson *et al.* 1989). The key Japanese economic agencies, the Bank of Japan, the Ministry of Finance, and (as it was known at the time) the Ministry of International Trade and Industry (MITI) had a clear “vision” of Japan as a modern industrialized nation.<sup>2</sup> Later, other Northeast Asian nations encouraged by the example of Japan pursued similar goals (Wade 1990).

In 1970 the Vice-Minister (bureaucratic head) of MITI in the 1960s and 1970s, Mr Ojimi, explained that:

If the Japanese economy had adopted the simple doctrine of free trade and had chosen to specialize in this kind of [labour-intensive] industry, it would almost permanently have been unable to break away from the Asian pattern of stagnation and poverty ... MITI and industry decided to establish in Japan industries which require intensive employment of capital and technology, industries that in consideration of comparative cost should be the most inappropriate for Japan, industries such as steel, oil refining, petrochemicals, automobiles, industrial machinery and electronics. From a short run static viewpoint, encouragement of such industries would seem in conflict with economic rationalism. But from a long range viewpoint, these are precisely the industries where income elasticity of demand is high, technological progress is rapid, and labour productivity rises fast.

(OECD 1970 cited in Eatwell 1987: 738)

Thus the key goal of industry policy was to support industries which had the following characteristics:

- 1 high income elasticity of demand;
- 2 considerable scope for productivity increases;
- 3 capability of contributing to an integrated industrial structure;
- 4 prospects for export.

MITI determined that industries with these characteristics were primarily in the manufacturing sector. Once agreement had been reached on the general goals of industry policy, a series of Plans was developed to implement the policy. These included the Five Year Plan for Economic Independence (1955–9), New Long Run Economic Plan (1958–62), and the Plan for Doubling National Income (1961–70) (Komiyah and Itoh 1988: 177).

### ***Second element: complementarity of import substitution and export-led growth***

Perhaps the most remarkable aspect of this initial development process is that it was largely based on production for the domestic market. The policy of import substitution was central to the creation of an integrated industrial structure.



Komiya and Itoh (1988: 186), referring to the first few post-war decades, bluntly state that, "it is not correct to characterize Japan's economic growth as export led." Yoshikawa (1993: 9) argues that "economic growth in this period [1950–70] was basically led by domestic demand." The primary evidence for the case that growth was based on domestic demand is that from the mid-1950s through to the late 1970s Japan had the lowest ratio of exports and imports to domestic production in the OECD outside of the US (Eatwell 1982: 139; Komiya and Itoh 1988: 175).

There were several factors enabling high rates of domestic-led growth to be sustained. These included the promotion of local industries, which generated high levels of demand for domestic output, major shifts in industry structure away from agriculture to manufacturing, and related changes in population centres and household formation.

The instruments of local industry promotion included tariffs, quotas, health and administrative regulations used to restrict imports and government procurement practices which ensured that domestic demand would be reserved for domestic production. The tariff regime had a cascading structure designed to restrict foreign competition with domestic production while minimizing tariffs on imports of essential raw materials and essential non-competitive manufactured products. Low tariffs were also set on industries deemed to have limited prospects for future development or, from the 1970s, on heavily polluting and energy-intensive industries which government sought to move offshore. The Ministry of Finance also played a major role in the very high levels of public sector infrastructure investment, such that the ratio of public fixed capital formation to GDP after the Second World War was two to three times higher than in the US, UK, France and Germany during the same period (Stein 1993: 112). Public infrastructure investments contributed significantly to domestic demand in the construction and manufacturing industries and complemented private investment capacity. Import substitution, however, was also combined with export promotion. While the period of rapid growth was marked by comparatively low export to GDP ratios, exports were nevertheless critical to overcoming Japan's chronic balance of payments deficit, which had been a feature of Japan's industrialization since it began in the mid-nineteenth century. As a resource-poor nation Japan needed to generate a level of exports sufficient to fund the import of essential raw materials necessary for the ambitious programme of industrialization. Policies to promote exports included subsidizing overseas sales by higher domestic prices and lower company tax on export income. Overseas market intelligence was also provided to firms, mainly through the Japanese External Trade Organisation (JETRO), the overseas arm of MITI (Eads and Yamamura 1988: 25).

The apparent paradox of high levels of simultaneous import protection and high levels of exports has been identified by Wade (1990) and others and was designed to accelerate the process of learning by doing/using and to realize increasing returns by increasing the size of the market available to firms. It is "misleading ... to present import substitution and export promotion as mutually exclusive strategies ... at the individual industry level, import substitution and export promotion can be complementary" (Wade 1990: 363). A key strategy for

preventing “x-inefficiency” in protected industries arising from the absence of international competition was the generally temporary nature of assistance or the tying of assistance to the achievement of specific performance measures such as productivity gains or export targets (Wade 1990: 359; World Bank 1993: 9).<sup>3</sup>

Another major cause (and effect) of increased domestic demand was population movement from rural areas to the cities associated with industrialization. These population movements simultaneously created the necessary labor force for manufacturing and large-scale demand for manufactures. The period of rapid growth was achieved without a large natural population increase or immigration; from 1950 to 1970 the population increased by only 1 per cent per annum. The labor force for the period of high growth was derived principally from surplus agricultural labor; in 1950 some 50 per cent of the workforce was employed in agriculture, only 22 percent of the workforce was in manufacturing. By 1970 only 19 per cent of the workforce was in agriculture and manufacturing’s share had risen to 34 per cent. A critical effect of this inter-industry labor mobility was that while the population increased by only 1 per cent per annum between 1950 and 1970, household formation increased by 80 per cent over the period. The rapid growth of urban households fed directly into the demand for consumer goods and infrastructure (Yoshikawa 1993).

The other advantage of being a late-comer to industrialization in this period was access to newly developed highly productive manufacturing technologies and product designs that were directly imported without the effort or expense of innovation (Stein 1993: 25–6).

### *Third element: selective industry policies*

The third element is the key role of the state through a variety of selective industry policy instruments to promote an integrated manufacturing base capable of propelling its own growth and development by coordinating investment and demand and exploiting increasing returns and external economies.

Johnson argues that industry policy operated on both “micro and macro aspects of the economy.” At the firm level industry policy involves:

state intrusion into the detailed operations of individual enterprises with measures intended to improve those operations ... in its simplest terms it is the attempt by the state to discover what it is individual enterprises are already doing to produce the greatest benefits for the least cost, and then ... to cause all enterprises of an industry to adopt these preferred procedures.

(Johnson 1982: 27)

At the industry or inter-industry level the concept of “industrial structure” was the key to planning by MITI. This planning:

concerns the proportions of agriculture, mining, manufacturing, and services in the nation’s total production; and within manufacturing it concerns the

percentages of light and heavy, and of labour-intensive and knowledge intensive industries. The application of the policy comes in the government's attempts to change these proportions in ways it deems advantageous to the nation. Industrial structure policy is based on such standards as income elasticity of demand, comparative costs of production, labour absorptive power, environmental concerns, investment effects on related industries, and export prospects.

(Johnson 1982: 28)

The instruments of industry policy were either direct or indirect. Direct instruments included tax concessions for firms to engage in prescribed activities, import quotas, incentives for export, government infrastructure projects and procurement contracted exclusively to domestic firms and control of firms' access to foreign currency and technology licences. The main indirect instrument was "administrative guidance" to banks and firms involving direction over the quantity of loan funds and industry sectors to receive these funds. Given the extensive controls exercised by the bureaucracy the system operated, in the words of Eads and Yamamura (1988: 433), on "rewards and retributions."

The post-war economic planners promoted a concentrated industry structure designed to exploit scale economies at an enterprise level and eradicate what MITI termed "excessive" or "wasteful" competition which it "feared ... would be associated with excess plant capacity, predatory pricing policies and low profit margins ... the government believed it had a duty to shape the country's industrial structure along lines similar to those of its leading competitors" (Stein 1993: 15; see also Aoki *et al.* 1996). It was acknowledged that the realization of scale economies at the level of enterprise requires limitation on the entry of firms to any industry. One of MITI's earliest acts was to dilute the provisions of the Anti-Monopoly Act, a strong anti-trust legislation imposed on the Diet by Occupation Forces in 1947. In 1949 and 1953, major provisions of the Act were repealed and, as enforcement of the Act was MITI's responsibility through the Office of Fair Trading, the Ministry ensured that there were very few prosecutions, and that penalties were very limited (Yamamura 1990: 38). Under conditions of falling average costs with market prices given, an individual firm can increase total profit by increasing output. "Additional market share pushes a firm down its cost curve, setting off a continuing cycle. As the firm increases volume, it takes additional market share, which lowers its costs, making it able to increase sales, thus starting the cycle over" (Tyson and Zysman 1989: 84; see also Wade 1990: 351–2). If the firm's increased capacity cannot be utilized because of intense rivalry with other firms, the cycle can be reversed. Policies to foster concentration were, however, tempered by strategies to promote competition between a few large firms or a few large networks of firms (*keiretsu*). Within the domestic market rivalry between several large firms in each industry (such as steel, automobiles and electronics) was encouraged and controlled (Stein 1993: 15; Murakami 1988: 54; Hart 1992: 78). In other words, "competition was bounded and orchestrated" (Johnson *et al.* 1989: 77).<sup>4</sup>

Rapid capital accumulation, concentration of production and wage restraint allowed firms to capture a large share of productivity gains and to reinvest these funds. Over the period of rapid growth from the 1950s to the early 1970s there was a large shift in national income from households to the corporate sector (World Bank 1993: 238). "One of the most intriguing aspects of the post war Japanese economy is what appears to be a reversal of economic roles. Consumers serve producers rather than vice versa" (Yamamura 1990: 40).

The objective of industry policy was to create an integrated industrial structure to maximize backward and forward linkages in a virtuous cycle of increasing output, incomes, investment, productivity and demand. The success of these policies is evident in the highly skewed composition of Japanese foreign trade which was marked by a very low level of intra-industry transactions, whereas trade across almost all OECD nations is marked by much higher levels of such transactions (Yamamura 1990: 43).

Wade (1990: 353) argues that "[a] big push, involving simultaneous expansion of several industries, can insure the profitability of each investment, even though each would be unprofitable on its own ... such simultaneous expansion helps to overcome the constraint of a small domestic market." An excellent example of the role of the state in redressing a series of vicious circles impeding development is provided by Okazaki (1996). In 1950s Japan, capital goods makers were keen to export, but the price of Japanese steel was high due to the small scale of production and the high cost of both local coal and shipping. (Most of Japan's merchant fleet had been destroyed in the war.) The interests of four industries were linked, namely machinery makers, steel producers, shipbuilders and local shipowners. The way out of this vicious circle was co-ordinated investment and consumption decisions across the different industries. The state subsidized a major expansion of capacity by steel producers, allowing them to achieve international scale economies. The state also encouraged major expansion of shipbuilding capacity which provided a large market for steel as well as reducing the cost of shipping, which in turn reduced the cost of imported raw materials for the steel industry. Machinery manufacturers benefited from both cheaper steel and shipping costs. Shipowners benefited from the large imports of raw materials and exports of machinery.

The common feature of industry policy was the targeting of specific industries over time to build up a chain of increased value-adding, technological sophistication and capital intensity. Japanese industry policy was a dynamic developmental strategy "critically concerned about the links between current resource allocation and the future evolution of the economy" (Johnson *et al.* 1989: 11). The MITI planners argued that the future growth and technological development of a nation is dependent on current and past industrial structure and composition of production. As a result of its economic history each nation is on a developmental trajectory. "If the dynamic potential of economic activities differs, then a national specialisation at a given moment, which is efficient in terms of current resource allocation, may not maximize economic welfare in the long run" (Johnson *et al.* 1989: 14).

In addition to the concentrated industry structure promoted by MITI, several authors emphasize the unique role of vertical and horizontal *keiretsu* within this industrial structure (Johnson 1982). Vertical *keiretsu* are formed by long-term commercial relations between manufacturers and subcontractors. The best-known vertical *keiretsu* were in the automobile and electronics industries, namely Toyota, Matsushita and Sony. These vertical *keiretsu* are a “balance between vertical integration and independence,” where the principal manufacturer provides financial and technical support to suppliers for the purpose of improving the quality and timeliness of component supplies (Yamamura 1990: 32). “Close links between assemblers and suppliers economize on transaction costs, enhance the transfer of technology, and increase incentives to make specific investments” (Lawrence 1993: 12). Horizontal *keiretsu*, of which the major ones were Mitsui, Mitsubishi, Sumitomo, Fuyo, Dai-ichi Kangyo and Sanwa, each comprised a few dozen firms including a bank, manufacturers and distribution companies. These companies are closely tied through high levels of cross shareholding and interlocking directorates, and they engage in inter-company financing, purchase and supply and joint R&D ventures (Gerlach 1989).

From the perspective of the circular and cumulative school, the *keiretsu* organization is explicable in that such an organization of economic activity creates and internalizes a number of positive externalities. Firstly, as emphasized by Rosenstein-Rodan (1943), the joint planning or close integration of investment projects may convert marginal investment projects into profitability and increase the scale of such projects. Coordination reduces the risk inherent to capital investment. Second, the role of distribution or trading companies in providing exclusive wholesale and retail outlets to *keiretsu* members and inhibiting the marketing of competitive imports facilitates industry concentration and exploitation of increasing returns (Yamamura 1990: 45).

### **Explaining the efficacy of industry policy in recent writings on the developmental state**

In 1999, Chalmers Johnson nominated a number of writers as having taken up the mantle of “the developmental state” (Johnson 1999: 35). These were Wade, Amsden (1989), Woo-Cumings and Yu-Shan. To these we may add Weiss, Krieckhaus, Pempel and Haggard, who also contributed to the discussion of Wade’s new introduction to his *Governing the Market* in a special issue of *Issues and Studies* in 2004.<sup>5</sup>

In his reflections on *Governing the Market* 15 years after its publication Wade emphasized the contribution that CC theory, had it been explicitly employed, could have made to the original book. “But *Governing the Market* says enough about the invisible strings between industrial policies and economic performance to make it plausible that the policies and their implementation agencies were too important in [North] East Asia’s success to ignore” (Wade [1990] 2003: xvii). In particular he noted the importance of linkages between industries within an economy.

We need to reintroduce a distinction that has dropped out of the development lexicon. The word “integration” is currently used to refer only to integration into the world economy, and carries with it the implication that more integration is always better. We should distinguish between “external integration” and “internal integration” (or articulation), and recognize that the development of a national economy is more about internal integration than about external integration.

(Wade [1990] 2003: xlviii)

Further,

[a]n economy with high internal integration has a well-filled input–output matrix – a dense set of links between sectors (a high level of sectoral articulation between, e.g., rural and urban, and consumer goods and intermediate goods), and a structure of demand such that a high proportion of domestic production is sold to domestic wage earners (a high level of “social” articulation between wages, consumption, and production).

Wade’s appeal was finally made somewhat confusing though by the following remark: “In short, the central challenge of national development strategy is to combine the principle of comparative advantage and the principle of import replacement” (Wade [1990] 2003: 1).

In his rejoinder to the various participants in the 2004 conversation about the new Introduction to *Governing the Market* (and especially to the present authors), Wade represented CC as a means to understanding the efficacy of selective industry policy. He wrote that “the purpose of state industrial policy is to exploit increasing returns, externalities, and complementarities so as to accelerate income growth and structural change” (Wade *et al.* 2004: 118). He combined this endorsement, though, with an entirely correct caution that CC has to be married with a political analysis of the development process.

Amsden is one of the few writers on Northeast Asian development to explicitly draw on the CC tradition and to argue that her empirical findings support these ideas. Amsden’s (1985) work on the development of Taiwan’s machine tool industry provides an important example of a detailed examination of the operation of increasing returns within a particular industry. Amsden’s historical case study approach focuses on long-term changes in the organization of production and technology within industry and on individual firms as they arise from growth in size of the market and the exploitation of increasing returns. Amsden stated that this method is essential “to try to understand the cumulative causation process involving fast economic growth in newly industrialising economies” (Amsden 1985: 282). Further, her findings support Kaldor’s catholic views on the sources of increasing returns (Amsden 1985: 273). In *Asia’s Next Giant* Amsden (1989) argued that the Korean state deliberately “got prices wrong” by paying conditional subsidies to selected *sectors* and through pressure on firms during negotiations with *chaebols* over controlled prices. Performance

conditions put upon subsidized industries were enforced by punishments and poorly performing firms could be “cold-bloodedly” amalgamated with other, more successful firms (Amsden 1989: 15). The Korean state was not averse to establishing public enterprises in preferred industries. The overall strategy of the Korean state was characterized as market augmenting, promoting a dynamic in which “not only does higher productivity generate higher growth ... [but] ... higher growth also generates higher productivity by means of learning-by-doing, economies of scale and investments embodying foreign designs” (Amsden 1989: 153). The acquisition of leading overseas technologies and “foreign designs” was directly sponsored by the state. More recently in *The Rise of the Rest* Amsden (2001) gave particular prominence to the role of developmental states in the evolving “knowledge economy.” She also extolled the importance of national firms and national innovation (and thus the national retention of technological rents), and of national formation of technical skills. The state targeted particular firms and, *through them*, particular industrial activities (Amsden 2001: 190).

Meredith Woo-Cumings in *The Developmental State* (1999: 27), who herself had contributed much on state-guided credit financing of industrialization, particularly in Korea, wrote approvingly of the “economic logic” of Ha-joon Chang (see e.g. Chang 2006) and no doubt would approve of Jomo’s arguments in the same vein (e.g. Jomo 2001, 2003). Analysts such as Jomo and Chang, from more of an economics background than a background in “international political economy,” tend to speak of the same group of circumstances of disequilibria as we have emphasized in our argument above (pervasive externalities, increasing returns to scale of enterprise and so on) and to emphasize the need for the state as coordinator of investment decisions in different industries. It is our argument, however, that they do not sufficiently recognize the mechanisms by which state industry policies could complement each other in promoting circular and cumulative causation. They tend not to see that state assistance of specific industries was selective and was arguably well designed to bring together the various factors contributing to the expansion of the manufacturing sector as a whole. Perhaps a little too colourfully one might complain that they fail to see the overall process as being one of harvesting external benefits and identifying the seedlings of new ones.

## Conclusion

This chapter has argued that the key elements identified by economic historians and political scientists such as those included in this survey regarding Northeast Asian post-war industry policy are strongly consistent with and supportive of the theory of circular and cumulative causation. We argue that the promotion of manufacturing by the key economic agencies was based on a conscious rejection of orthodox development prescriptions and that the success of industry policy is to be found in the exploitation of elements such as positive externalities, increasing returns and complementarities in production and investment. Such



elements are seen as inherent to the operation of industrial economies and are represented broadly as operating jointly in a cumulative process: they contribute jointly to the dynamic of economic growth. The purpose of state industry policies is to exploit these elements. For example, and perhaps most suggestively, state coordination of investments in neighboring industries (or *some* extra-market coordination) is necessary for the “harvesting” of positive externalities. The “statist” writers have concluded as well that strategies of the developmental states also involved simultaneous import substitution and export promotion within a single industry and across industries. These findings are important because they support the suggestions made by the leading CC theorists that Northeast Asian post-war growth is explicable from a CC viewpoint.

This study has also argued that a clearly articulated theoretical understanding of CC is generally absent not only from the work of political scientists but of some very significant writers with more of an economics background. There is thus a tendency to argue in terms of categories that really *only* make sense within a neoclassical perspective and thus to be confusing in their discussion of industry policies. The category of market failure is the clearest but by no means the only example. The creation of positive outcomes that are not appropriable by the agents whose activities give rise to them is inherent to industrial economies. The policy challenge is to exploit them. It is misleading to see them as “market failures” and to pose some sort of correction of such “failures” as the significant policy challenge. Even Robert Wade only very obliquely linked his own “Governed Market” model with the CC tradition.

There is a reasonable question as to whether the Northeast Asian developmental states of the 1970s and 1980s are evident still today; and there is a further question as to whether they can be emulated today by other countries, particularly in Southeast Asia. As to the first question, the collection of papers published in Volume 40 of *Issues and Policies* (Wade *et al.* 2004) provides various arguments to the effect that the older developmental states are indeed still working as such, although a little more subtly. Moreover, it is arguable that there is occurring a regionalization of the Japanese developmental state. Finally, the recent developmental state literature that has been cited in Part 2 above does put the argument that developmental states do exist in Southeast Asia, albeit that various of their capacities have been corrupted or have yet to be formed.

The very motivation of the CC School was to resist the apriorism of equilibrium economics. Without going overboard in the direction of empiricism, there is, on the other hand, room for a fruitful union of the detailed empirical investigations of development within particular national economies and the CC theory of dynamics in industrial economies and strategy of development. Such an approach would be consistent with the view of Allyn Young (1929), the originator of CC theory, who argued for the complementarity and mutual support of “economic theory” and economic history. Given the arguments presented in this chapter, one can certainly concur with Robert Wade’s assessment that there is substantial “reason for economists to accept the challenge of constructing a theoretical rationale for the non-neoclassical East Asian facts” (Wade 1990: 381).



## Notes

- 1 The voluminous literature on the Verdoorn law since Nicholas Kaldor's (1966) initial statement of the law attests to the particular empirical methodology adopted by the CC school. See e.g. McCombie *et al.* 2002.
- 2 In the immediate post-war years however, there was a major debate within the bureaucracy over development policy. In particular, the Governor of the Bank of Japan, the Prime Minister and leading academics favored a policy of "Tradeism" or integration of Japan into the world economy on the basis of comparative advantage. Japan would utilize its surplus agricultural labor force to concentrate on labor-intensive products such as textiles, clothing, pottery and small metal wares. Another group favored "Developmentalism," or the creation of capital intensive industries, believing that market-determined outcomes would permanently consign Japan to third world status. By the late 1940s, the "Developmentalists" had won a decisive victory with the establishment of the Ministry of International Trade and Industry in May 1949. In 1953 a systematic plan for establishing a diversified industrial structure with the focus on the growth of the domestic market was accepted by the Japanese Cabinet (Johnson 1982, 228–30; Stein 1993, 12; Johnson *et al.* 1989, 65; Allen 1981).
- 3 A key element in MITI's original development Plan in the mid-1950s was the expansion of the domestic market to facilitate export promotion.

Ishibashi [one of the chief architects of industry policy] pointed out that the key to exports was, of course, the lowering of costs, and the key to that was enlarging production to effect economies of scale. But to enlarge production, Japanese manufacturers needed more customers. And where were they to be found? In the huge potential market of Japan itself.

(Johnson 1982: 229)

- 4 Considerable attention was also paid to the small business sector. The 1963 Small and Medium Enterprise Law had a number of objectives which included *inter alia* "1. Modernisation of Equipment, 2. Improvement of Technology, 3. Rationalisation of Management, 4. Upgrading of structure (optimisation of business scale), 5. *Prevention of Excessive Competition* ... 6. Promotion of Exports and other demands" (Japan External Trade Organisation 1986: 11; italics added).
- 5 This special issue was edited by Andrew Marble and also included in the conversation with Wade Elizabeth Thurbon, co-writing with Weiss, and ourselves. See Wade 2003; Wade *et al.* 2004; Toner and Butler 2004.

## References

- Agarwala, A. N. and Singh, S. P. (1973) *The Economics of Underdevelopment*, New Delhi: Oxford University Press.
- Allen, G. C. (1981) *The Japanese Economy*, London: Weidenfeld & Nicolson.
- Amsden, A. (1985) "The Division of Labour is Limited by the Rate of Growth of the Market: The Taiwan Machine Tool Industry in the 1970's," *Cambridge Journal of Economics*, 9: 271–84.
- (1989) *Asia's Next Giant: South Korea and late industrialization*, New York: Oxford University Press.
- (2001) *The Rise of the Rest: Challenges to the West From Late-industrializing Economies*, New York: Oxford University Press.
- Aoki, M. (1988) "The Japanese Firm in Transition," in T. Inoguchi and D. Okimoto (eds) *The Political Economy of Japan*, Stanford, CA: Stanford University Press.

- Aoki, M., Hyung-Ki, K. and Okuno-Fujiwara, M. (eds) (1996) *The Role of Government in East Asian Economic Development*, Oxford: Clarendon Press.
- Aoki, M., Murdock, K. and Okuno-Fujiwara, M. (1996) "Beyond *The East Asian Miracle*: Introducing the Market-enhancing View," Chapter 1 in M. Aoki *et al.* (eds).
- Argy, V. and Stein, L. (1997) *The Japanese Economy*, London: Macmillan.
- Argyrous, G. (1996) "Cumulative Causation and Industrial Evolution: Kaldor's Four Stages of Industrialisation as an Evolutionary Model," *Journal of Economic Issues*, 30, 1: 97–119.
- Chang Ha-joon (2006) *The East Asian Development Experience: The Miracle, the Crisis and the Future*, London and New York: Zed Books, with TWN, Malaysia.
- Chenery, H. (1959) "The Interdependence of Investment Decisions," in M. Abramovitz (ed.) *The Allocation of Economic Resources*, Stanford, CA: Stanford University Press; reprinted in H. Chenery (ed.) (1979) *Structural Change and Development Policy*, Washington, DC: Oxford University Press, A World Bank Publication.
- Cornwall, J. (1977) *Modern Capitalism. Its Growth and Transformation*, Oxford: Martin Robertson.
- Deraniyagala, S. and Fine, B. (2001) "New Trade Theory Versus Old Trade Policy: A Continuing Enigma," *Cambridge Journal of Economics*, 25: 809–25.
- Dosi, G., Tyson, L. and Zysman, J. (1989) "Trade, Technologies and Development. A Framework for Discussing Japan," in C. Johnson, L. Tyson and J. Zysman (eds) *Politics and Productivity. The Real Story of Why Japan Works*, New York: Harper Business.
- Eads, G. and Yamamura, K. (1988) "The Future of Industrial policy," in T. Inoguchi and D. Okimoto (eds) *The Political Economy of Japan*, Stanford, CA: Stanford University Press.
- Eatwell, J. (1982) *Whatever Happened to Britain? The Economics of Decline*, London: Duckworth.
- (1987) "Import Substitution and Export-led Growth," in J. Eatwell, M. Milgate and P. Newman (eds) *The New Palgrave* (4 Vols), London: Macmillan.
- Gerlach, M. (1989) "Kiretsu Organisation in the Japanese Economy," in C. Johnson, L. Tyson and J. Zysman (eds) *Politics and Productivity. The Real Story of Why Japan Works*, New York: Harper Business.
- Gold, B. (1981) "Changing Perspectives on Size, Scale and Returns: An Interpretative Essay," *Journal of Economic Literature*, 19: 5–33.
- Hart, J. A. (1992) *Rival Capitalists. International Competitiveness in the United States, Japan, and Western Europe*, Ithaca, NY: Cornell University Press.
- Hatch, W. and Yamamura, K. (1996) *Asia in Japan's Embrace: Building a Regional Production Alliance*, Cambridge: Cambridge University Press.
- Hippel, von E. (2005) *Democratizing Innovation*, Cambridge MA: MIT Press.
- Hirschman, A. (1958), *The Strategy of Economic Development*, New Haven, CT, and London: Yale University Press.
- Inoguchi, T. and Okimoto, D. (1988) *The Political Economy of Japan*, Stanford, CA: Stanford University Press.
- Japan Economic Institute (1991) *Industrial Policies in Search of a Model: Comparing the Experiences of Japan and the United States*, Washington, DC.
- Japan External Trade Organisation (1986) *Japan's Postwar Small and Medium Enterprise Policy. For Promotion of Corporate Vitality*, Tokyo (March).
- Johnson, C. (1982) *MITI and the Japanese Miracle*, Stanford, CA: Stanford University Press.
- (1999) "The Development State: Odyssey of a Concept," in M. Woo-Cumings (ed.) *The Development State*, Ithaca, NY, and London: Cornell University Press.

- Johnson, C., Tyson, L. and Zysman, J. (eds) (1989) *Politics and Productivity. The Real Story of Why Japan Works*, New York: Harper Business.
- Jomo, K. S. (2001) *Growth After the Asian Crisis: What Remains of the East Asian Model?*, UNCTAD G-24 Discussion Paper series, 10, UN and Harvard Center for International Development.
- (ed.) (2005) *Southeast Asian Paper Tigers: From Miracle to Debacle and Beyond*, London and New York: Routledge Curzon.
- Kaldor, N. (1966) *Causes of the Slow Rate of Economic Growth of the United Kingdom*, Cambridge: Cambridge University Press.
- (1967) *Strategic Factors in Economic Development. The Frank W. Pierce Memorial Lecture*, Ithaca, NY: Cornell University Press.
- (1972) "The Irrelevance of Equilibrium Economics," *Economic Journal*, 82: 1237–55.
- (1978) *Further Essays on Applied Economics*, London: Duckworth.
- Komiya, R. and Itoh, M. (1988) "Japan's International Trade and Trade Policy 1955–1984," in T. Inoguchi and D. Okimoto (eds) *The Political Economy of Japan*, Stanford, CA: Stanford University Press.
- Lawrence, R. J. (1993) "Japan's Different Trade Regime: An Analysis with Particular Reference to Keiretsu," *Journal Of Economic Perspectives*, 7: 3–20.
- Leontief, W. (1963) "The Structure of Development," *Scientific American*, 209: 148–66; reprinted in *Input–Output Economics* (2nd edn 1986), Oxford: Oxford University Press.
- McCombie, J. and Thirlwall, A. P. (1994) *The Balance of Payments Constraint to Growth*, New York: St Martin's Press.
- McCombie, J., Pugno, M. and Soro, B. (2002) *Productivity Growth and Economic Performance: Essays on Verdoorn's Law*, London: Palgrave-Macmillan.
- Meade, J. E. (1952) "External Economies and Diseconomies in a Competitive Situation," *Economic Journal*, 62: 54–7.
- Murakami, Y. (1988) "The Japanese Model of Political Economy," in T. Inoguchi and D. Okimoto (eds) *The Political Economy of Japan*, Stanford, CA: Stanford University Press.
- Myrdal, G. (1958) *Economic Theory and Underdeveloped Regions*, London: Methuen.
- Nelson, R. R. and Pack, H. (1999) "The Asian Miracle and Modern Growth Theory," *The Economic Journal*, 109: 416–36.
- OECD (1970) *Japanese Industry Policy*, Paris.
- Okazaki, T. (1966) "Relationship between Government and Firm in the Post WWII Economic Recovery: Resolving the Coordination Failure by Coordination in Industrial Rationalization," in M. Aoki, M. Okuno-Fujiwara and H. Kim (eds) *The Role of Government in East Asian Development: Comparative Institutional Analysis*, Oxford: Oxford University Press.
- Pasinetti, L. (1981) *Structural Change and Economic Growth. A Theoretical Essay on the Dynamics of the Wealth of Nations*, Cambridge: Cambridge University Press.
- Romer, P. (1994) "The Origins of Endogenous Growth," *Journal of Economic Perspectives*, 8, 1: 3–22.
- Rosenberg, N. (1982) *Inside the Black Box: Technology and Economics*, Cambridge: Cambridge University Press.
- Rosenstein-Rodan, P. N. (1943) "Problems of Industrialisation in Eastern and South Eastern Europe," *Economic Journal*, 58, in A. N. Agarwala and S. P. Singh (eds) (1973) *The Economics of Underdevelopment*, New Delhi: Oxford University Press.
- Saxonhouse, G. (1988) "Comparative Advantage, Structural Adaptation and Japanese

- Performance," in T. Inoguchi and D. Okimoto (eds) *The Political Economy of Japan*, Stanford, CA: Stanford University Press.
- Scitovsky, T. (1954) "Two Concepts of External Economies," *Journal of Political Economy*, 62: 143–51; reprinted in A. N. Agarwala and S. P. Singh (eds) (1973) *The Economics of Underdevelopment*, New Delhi: Oxford University Press.
- Stein, L. (1993) *Japanese Industrial Policy*, Centre for Japanese Economic Studies Macquarie University, Working Paper 93–7.
- Thirlwall, A. P. (1979) "The Balance of Payments Constraint as an Explanation of International Growth Rate Differences," *Banca Nazionale Del Lavoro Quarterly Review*, 128: 45–53.
- (1983) "A Plain Man's Guide to Kaldor's Growth Laws," *Journal of Post-Keynesian Economics*, 4, 3: 345–58.
- Toner, P. (1999) *Main Currents in the Theory of Circular and Cumulative Causation: The Dynamics of Growth and Development*, Basingstoke: Macmillan.
- Toner, P. and Butler, G. (2004) "Some Reflections on Governing the Market," *Issues and Studies*, 40, 1: 81–102.
- Tyson, L. and Zysman, J. (1989) "Developmental Strategy and Production Innovation in Japan," in C. Johnson, L. Tyson and J. Zysman (eds) *Politics and Productivity. The Real Story of Why Japan Works*, New York: Harper Business.
- Verdoorn, P. J. (1949) "Factors that Determine the Growth of Labour Productivity," *L'industria* (trans. G. and A. P. Thirlwall; privately circulated).
- Wade, R. (1990; 2nd edn 2003) *Governing the Market: Economic Theory and the Role of Government in East Asian Industrialisation*, Princeton, NJ, and London: Princeton University Press.
- Wade, R., Haggard, S., Krieckhaus, J., Weiss, L., Thurbon, E., Pempel, T. J., Toner, P. and Butler, G. (2004) The Reprinting of *Governing the Market*, special number of *Issues and Studies*, 40, 1.
- Woo-Cumings, M. (Woo Jung-en) (1991) *The Rise of the Swift: State and Finance in Korean Industrialization*, New York: Columbia University Press.
- (1999) "Introduction: Chalmers Johnson and the Politics of Nationalism and Development," in M. Woo-Cumings (ed.) *The Developmental State*, Ithaca, NJ, and London: Cornell University Press.
- World Bank (1991) *World Development Report 1991: The Challenge of Development*, New York: Oxford University Press.
- (1993) *The East Asian Economic Miracle. Economic Growth and Public Policy*, New York: Oxford University Press.
- Yamamura, K. (ed.) (1990) *Japan's Economic Structure: Should it Change*, Seattle: Society for Japanese Studies.
- Yoshikawa, H. (1993) "High Economic Growth of the Japanese Economy and Its End: An Explanation by a Model of Demand-led Growth," Macquarie University Conference on the Contemporary Japanese Economy.
- Young, A. (1928) "Increasing Returns and Economic Progress," *Economic Journal*, 38: 527–42.
- (1929) "Economics," in *The Encyclopaedia Britannica*, 14th edn, Vol. 7: 925–32.
- Young, Alwyn (1995) "The Tyranny of Numbers: Confronting the Statistical Realities of the East Asian Growth Experience," *The Quarterly Journal of Economics*, 110, 3: 641–80.

# 4 Cumulative causation and industrial development

## The regional stage

*George Argyrous and Geoff Bamberry*

### Introduction

Nicholas Kaldor and Gunnar Myrdal are regularly cited as central figures in the development of the theory of circular and cumulative causation. They are seen as complementary figures, each improving our understanding of different aspects of the broader circular and cumulative causation process. Kaldor ([1966] 1978: 198) emphasized the role of manufacturing and increasing returns to scale as the engine of growth, and the possible limiting factor of effective demand, while Myrdal (1957) emphasized the importance of socio-economic factors in the development process.

There does exist, however, a fundamental methodological difference between these two key figures regarding the role conceptual stages of growth play in understanding the process of circular and cumulative causation. This difference centres on the extent to which we can employ distinct “stages of growth” in our conceptual understanding of the circular and cumulative causation process.

Myrdal (1968), on the one hand, explicitly rejected the use of “stages” for explaining historical development. His most explicit treatment of this issue is in Appendix 2 of *Asian Drama*. There he gave four reasons why stages theories were not compatible with the theory of circular and cumulative causation:

- implicit in stages models is a belief that the transition from “early” to “late” stages is inevitable;
- there is a bias against the need for government intervention through active policy to give shape and direction to the development process;
- they cannot explain events that do not fit the preconceived scheme of stages;
- in the face of conflicting evidence, the evidence is dismissed by qualifications and reservations that make the stages approach tautological.

It is of interest to note that he was writing at a time when Rostow’s *Stages of Economic Growth*, and its conservative political implications, was particularly influential; Myrdal (1968) constantly directs most of his general criticisms of stages theories specifically to Rostow, but in so doing we believe he unfairly criticizes *all* stages theories.

Kaldor ([1966] 1978: 113), on the other hand, explicitly saw cumulative causation occurring within distinct stages, although he was not always consistent in his demarcation of these stages, and the processes that drive the system from one stage to another. He observed that market development by manufacturers occurred through a four-stage process. In the first stage firms meet the domestic demand for products, often providing a substitute for goods previously imported. The source of increased consumer demand is a rise in real income per head of population, which results in more disposable income being available for the purchase of additional manufactured products. While demand can increase for a time through this process, it eventually flattens out, limited by the size of the domestic market. At this point the cumulative effect begins to falter, and can turn into a vicious circle of decline unless demand can be increased from other sources.

In the second stage, manufacturers start to meet demand from overseas markets while continuing to supply national markets. The economies of scale achieved in the national market provide a step-up into international markets (Kaldor [1966] 1978: 113–14; [1981] 1989: 204; Targetti 1992: 177; Argyrous and Sethi 1996: 486). This provides a feedback effect to local capital goods manufacturers who begin a process of import substitution, which thereby increases the ability of local consumer goods manufacturers to make the transition to exporting (Kaldor 1972: 125; Thirlwall 1987: 324; Toner 1999: 125). In the third of Kaldor's stages, the production of capital goods for use by local mass production firms becomes firmly established, and leads to increasing specialization in the production of capital goods.

Finally, in the fourth of Kaldor's stages, the capital goods producers develop the expertise and scale of operation to a point where they can begin exporting. Over time, this move through stages of development allows manufacturing processes to be broken down into sequential steps through specialization and division of labour. In time, some processes are split off as separate enterprises, resulting in linkages between vertically integrated industries in the supply chain (Bamberry 2006a; Roberts 2000: 38; Toner 2000: 28).

Kaldor's articulation and application of his stages theory of circular and cumulative causation avoids Myrdal's concerns that we listed above. Kaldor did not envisage development as inevitably leading to a transition from one stage to the next; indeed he saw the stages as specific junctures in the industrialization process where the process of circular and cumulative causation could break down unless government policy was favourable to the transition. Moreover, his concern with the impact that aggregate effective demand could have on more localized processes of circular and cumulative causation avoided much of the potential teleological implications of taking a staged approach. For example, Argyrous ([1995] (1996)) has used Kaldor's four-stage model to explain both the growth and limits to the growth of the Australian machine tools industry.

The methodological task for proponents of circular and cumulative causation theory is to ensure that the right type and number of stages have been conceptually defined as a guide to empirical analysis of actual historical processes. Too

few stages can render the empirical analysis mere story-telling of detailed and case-specific events, while too many stages can blur the distinction between processes that drive a system within its established boundaries, and those that allow it to cross boundaries and move into the next stage of development.

In this chapter we take Kaldor's model one stage further; that is, we argue that he did not allow for localized, regional development as a starting point for a process that may eventually lead to international trade. While he drew attention to the potential for development at the regional level to create competition that would disadvantage weaker regions (Kaldor 1970: 144), he implicitly assumed that firms begin operating within the national market. By inserting an earlier stage whereby firms initially target their regional market, we gain a deeper understanding of the industrialization process, which provides a new dimension to possible government policies that facilitate this process. We argue that the delineation of stages of growth is marked by points in a firm's development where significant changes in the type and scale of capital goods investment needs to take place for economies of scale to be realized and the circular and cumulative process to continue. We illustrate this with reference to the development of the Australian wine industry. This case study provides an illustration of two more general points we wish to make:

- (1) Regional demand is an important early stage for industries with a significant agricultural base, because meeting regional demand does not rely on extensive capital investment, especially in terms of plant and equipment, storage, transport and distribution channels.
- (2) Regional demand is significant for capital goods producers, as the interaction between the producers and users of capital goods requires a close working relationship essential to the learning process involved in capital goods development.

## **Research methodology**

Cumulative causation is an essentially dynamic and contextual process that is given direction by specific historical circumstances. To understand this process, and specifically to determine the appropriate delineation of stages of development, qualitative methods of data collection, unfamiliar to many economists, are particularly useful (Ticehurst and Veal 2000: 95; Kerlinger and Lee 2000: 589). The empirical information upon which this chapter is based comes from 11 case studies using in-depth interviews, and in some cases, site inspections, to collect information. The study focused on a particular industry – the wine industry – and the interviewees were owners and/or managers of small and large wineries based in two of the main wine-growing regions of Australia: the Riverina and Hunter Valley Regions of New South Wales.

Firms were selected on the basis that they had been operating for at least six years, were engaged in sales in at least regional and national markets, preferably with some level of exporting or having considered the possibility of future



exporting. These enterprises were selected so that the process of industrialization across export stages could be adequately explored. The interview schedule for these interviews incorporated questions regarding the ownership, history, size and operations of the wineries. The interviews, undertaken in the latter part of 2005 and early 2006, lasted for approximately one hour, and each interview was recorded and transcribed for further analysis. Where necessary, statistical information obtained from interviews has been updated with data from the firms' websites.

In addition, we draw upon a case study of a capital goods firm directly related to the wine industry which was undertaken for earlier research (Bamberry 2006b). This involved an interview with a manager/director of A&G Engineering, a firm frequently referred to by winery owners and managers in the Riverina Region as being one of their most significant suppliers of capital equipment. We draw upon this material to illustrate the interaction between the producers of consumer goods (in this case wine) and capital goods that is central to Kaldor's stages of circular and cumulative causation.

The wine industry was chosen because we believe that the regional stage of development is particularly important to a class of industry that relies on a specific agricultural input (in this instance wine grapes). As we will discuss below, the reliance on an agricultural input, even where a proportion of value is subsequently added through manufacturing processes, means that firms in such an industry cannot directly launch into national markets, due largely to the changes in the capital goods investment required to take the step from regional to national markets.

## **Regional markets**

The interviews of the wine growers make it clear that regional markets are critical in the early stage of firm growth. Most of the firms in the study started off supplying regional markets, generally by establishing cellar-door sales outlets on the site of the combined vineyard and winery. This required a very limited amount of capital investment to tap local consumer demand. Many small wineries on popular tourist routes are able to sell most of their wine through cellar-door sales at a higher price than they can achieve through wider national and international markets. Where demand at cellar-door sales is insufficient to sell all the output, this is often supplemented by sales to local restaurants, hotels and locally owned liquor shops, all of which can be supplied with firm's existing transport facilities.

An example of a winery that started in the regional market before expanding nationally and internationally is Westend Estate Wines, established on the outskirts of Griffith in 1945 by immigrants from Italy, and taken over in 1974 by the founders' son who is the company's chief wine maker. Local and regional sales were the starting point for the firm, and in its 60-year history, this market has been developed incrementally using a number of strategies, the first of which has been the continuous improvement over time of the firm's cellar-door sales.

Like many wineries, these would have developed from sales at the door of the winery, often a corrugated iron shed, through many gradual improvements, to the purpose-built high-quality retail outlet of today (Westend Estate, webpage).

Some of the smaller wineries that are close to but not directly on major tourist routes in the Hunter Valley, such as Molly Morgan Wines and Glendonbrook Wines, have established small-scale high-grade tourist accommodation to take advantage of their location, landscape and history to attract customers, and to generate additional income (see webpages for photographs and details). A small family-owned winery, Broke's Promise Wines, has taken advantage of an olive grove attached to their vineyard, together with family connections in art and sculpture, to establish a retail outlet for olive-based products, as well as an art gallery and café in conjunction with expanded cellar-door sales facilities (see webpage). For these firms, increased demand has been sought through a strategy of seeing themselves as part of the tourism industry as well as part of the wine industry.

Previous research indicated that many firms have been able to sell most or all of their respective production in the regional market, so that there has been little apparent incentive to meet demand beyond the regional level (Wickramasekera and Bamberry 2001, 2003). The more recent research has confirmed this; however, in the course of meeting this regional demand, firms incrementally changed their methods of production to create the *capacity* to enter larger markets. This included upgrading of wine-making skills by sponsoring the education of family members in university and college wine-making courses, by employing graduate wine makers, or by contracting out production to experienced wine makers in neighbouring wineries. It also included the development of new or improved laboratory facilities and recruitment of research staff, often on a small scale, but significant in terms of the quality of the wine produced.

### **Moving beyond the regional stage**

The previous section emphasized that the wine-producing firms began with an almost singular focus on local sales, with incremental changes to operations aimed at meeting demand from regional sources. The interviews made it clear that this local stage was thought necessary, even where the long-term strategy of the firm was to eventually enter national and international markets. The interviews also showed that in the minds of the firms' owners, the transition from regional to national markets represented a distinct change in methods of operation and capital investment.

This transition for the individual firm from regional to national and export markets is itself affected by the extent to which other firms in the industry have already made the transition. That is, the level of development of the whole industry affects the pace at which individual firms can move from the local to the national and international stages of development. This is an important element of the cumulative causation story; the ability for a particular industry to expand is partially dependent on its own past success.

The development process in the wine industry, for example, was generally slower for wineries that commenced operating more than 30 years ago. As the “pioneers” of the national wine market, these older firms had to engage in a process of market creation at the national level, which more recent firms could then take as given when moving from the regional to the national level. Until the 1980s, the quality of wine was generally well below today’s standards, and the amount of wine consumed at the national level was below current per capita consumption levels. Limited demand, combined with the limited supply and relatively low prices, tended to keep the smaller wineries operating on a small scale for very long periods before development had proceeded to a point where expansion beyond the regional level could be undertaken.

However, over time, this situation changed, with the larger wineries improving the quality of their wines and attracting a larger consumer base. Consumers began to appreciate a range of varieties and styles, creating niches that newly established wineries could fill to begin their own expansion path. At the same time, technological improvements were occurring such as the introduction of stainless-steel vats and equipment to replace old wax and concrete vats that hindered improvements in wine hygiene and quality. Similarly, as some firms expanded into the national market, the introduction of viticulture and wine science courses at tertiary institutions further improved the quality and diversity of wines, making it easier for newly established firms to make the transition from regional to national markets.

While the move from the regional to more geographically dispersed markets was pursued through various strategies at the firm level, the “tipping point” at which a firm decided to begin the process of entering national markets was frequently cited as an anticipated oversupply of wine that could not be absorbed by local sales. Initial sales in markets outside the local region often involved selling through restaurants and specialized wine shops in larger urban centres, later moving into some of the smaller retail chains that were prepared to take limited stocks of higher priced “new” wines. Other steps for these firms in the process of expanding markets included:

- Improving quality through better hygiene by moving away from concrete vats to high-grade stainless-steel equipment.
- Implementing procedures to ensure appropriate temperature control in the processing and storage of wine.
- Establishing laboratories to test the wine at various stages of processing.
- Improving the marketing of wine by entering national and international wine competitions, winning medals and increasing the reputation of the wines produced.
- Selling wine in larger volume casks. For example, Baratto Winery, near Griffith, established over 40 years ago, and now operated by the original owner’s son, considered its current size and location not warranting capital expenditure on major cellar-door sales facilities. The owner decided on a strategy of “value for money” cask wines, sold mainly by word of mouth

through mail and telephone orders to customers in urban centres throughout Australia.

- Establishing various channels of marketing and sales, such as word of mouth, mail and telephone orders, and the use of specialist wine brokers. For example, Terrel Estate Winery, owned by a Spanish immigrant family who established a vineyard to sell grapes to wineries in the Griffith area, purchased a winery when a production glut reduced prices. The firm made use of wine brokers owing to its lack of experience in wine marketing. Having established a good working relationship with brokers in the domestic market, it was a logical progression to use their expertise to enter the export market. This strategy allowed the firm to sell wine in bulk to the brokers, who handled the transport, bottling and sales, as well as taking the risk in exporting.
- Improving economies of scale as demand increased, including moving from simple bottling lines to more complex and faster ones. Terrel Estate Winery was able to take advantage of its early success in export markets to purchase a relatively new bottling line from Casella when the latter firm upgraded its plant to further expand production.
- Inter-firm cooperation. For example, Penmara Wines was established by a group of five small to medium-sized wineries as a joint venture to market their wines in national and international markets while maintaining their separate identities, and continuing sales under their own brands in regional and national markets. The joint venture produces a range of wines under the Penmara label, focusing on particular demands in the domestic and international markets (see webpage for examples). As well as seeking to achieve economies of scale by producing quantities that the individual wineries might not be able to achieve, Penmara also provides scale economies in marketing.
- Product differentiation, whereby firms produce both high-priced quality wines with distinctive brand names, as well as lower priced brands. This is a strategy used by Penmara, the joint-venture firm, and larger firms such as De Bortoli, Casella and Westend who have the scale economies to produce lower priced wines in volume (see webpages for examples).

These general points are illustrated by the specific histories of some of the firms in the case studies. For example, De Bortoli Wines, established near Griffith by Italian immigrants in 1928, and now managed by the second and third generation of the founders' family, began its entry into the export market some years ago by selling wine in bulk to exporters who sold it under their own labels. Recognizing the disadvantage of not establishing its own brand name, the firm withdrew from this arrangement and later entered the export market under its own name. Other strategies pursued in the domestic market included expanding its original cellar-door outlet, purchasing existing wineries in tourism-oriented regions including the Hunter, King and Yarra Valleys, and establishing other attractions such as function centres and restaurants at the newly purchased wineries (see webpage for photographs).

As a result of its success, by 2005/6, the company was able to undertake an \$84 million programme to increase its crushing capacity by 80 per cent, adding 65,000 tonnes per year to its existing production capacity of 85,000 tonnes. In addition, the expansion included a new high-speed bottling line, extra warehousing, increased wine storage, as well as environmental measures to manage odour emissions, groundwater and water storage, and the impact of noise and dust. The numerous million litre stainless-steel storage tanks were augmented by new state-of-the art three-million-litre storage tanks. Here we see the interaction between consumer goods production and capital goods production that is central to Kaldor's stages of circular and cumulative causation, which we will discuss in more detail below.

The experience of Casella wines similarly illustrates the point that the move from one of the stages of development to another is marked by a significant increase in the production capacity of the firm. Established by Italian immigrants who bought a farm near Griffith in 1965 and set up a winery in 1969, the firm, now a very large exporter, is managed by the second and third generations of the family. Increased demand for the firm's wines following a major expansion into export markets resulted in the replacement of bottling lines, which were only a few years old, with new, high-volume, high-speed lines (Casella, webpage).

Other technological developments at Casella include the installation of five new presses, three centrifuges and over 60 million litres of storage capacity, making the winery capable of crushing 120,000 tonnes of grapes during a vintage. The firm also installed three bottling lines with a combined output of over 30,000 bottles per hour, with the intention of adding two more lines to give the winery a bottling capacity of over 65,000 bottles per hour. Staff numbers in the winery were also increased to over 320. Through this capital development and by using grape varieties that would appeal to a wide range of tastes, export sales worldwide grew from 500,000 cases in 2001 to almost 11 million cases (132 million bottles) in 2006/7 (Casella, webpage).

Casella has also sought scale economies by undertaking capital expenditure on environmental improvements, building the largest wastewater treatment scheme of its type in Australia at the time. All wastewater from the winery is recycled, enabling the firm to use an average of 2.5 litres of water for every one litre of wine produced, compared with the industry average of approximately 3.5 litres per litre of wine. The estimated annual costs of about \$150,000 to \$200,000 to treat the wastewater is equivalent to 38 to 50 cents per kilolitre, which is cost-effective as well as being environmentally sound (Casella, webpage).

## **Capital goods and local demand**

We have seen that localized markets are significant in the development of firms that may eventually develop an export capacity. In the course of moving from the regional to the national and export markets, producers of consumer goods increase their own demand for specialized capital equipment, and this creates a new regional source for industrial development. That is, the emergence and growth of a domestic capital goods sector depends on the highly localized demand for capital

goods that comes from producers of consumer goods. This is due to the degree of interaction that must occur between the producer and user of capital equipment. Capital equipment is often custom-built to meet the needs of individual production lines, and this becomes even more pronounced as the producer of finished goods adopts mass production technology (Argyrous [1995] 1996).

Scott and Storper (1992: 14) argue that “technical innovations are often place bound”, as “the stocks of human knowledge and human capital upon which technological changes are based tend to be concentrated in the specialised labour forces which themselves are highly localised”. Martin (1999: 79–80) also comments on the significance of location in technological development, referring to “geographically constrained interdependence such as technological spillovers, knowledge circulation, technical know-how and learning among the workforce”.

The wine-producing firms were asked about their relationships with equipment suppliers, and the results are typical of this process of interaction and learning at the local level. This is illustrated by the example of A&G Engineering. This firm was founded by Ron Potter in 1963 while working as a wine maker for a small winery. Through collaboration with the local wine industry the firm built up its production of specialized capital goods to meet the particular needs of these producers. This included new fermentation machinery, as well as a range of stainless-steel equipment and storage tanks (see A&G Engineering webpage). While building up the demands for its capital equipment in the early years of the firm’s development, the company achieved viability by becoming the local agent for Chamberlain farm machinery, as well as undertaking general engineering work for farms, wineries and other industries in the local area. As the firm’s reputation for its specialized capital equipment grew, it became less dependent on the agency role and on the local region for its markets, gradually expanding into national and international markets.

Working in collaboration with the local wine industry resulted in finding a solution to the problem of removing sulphur dioxide from bulk-stored wine through the invention of the spinning cone technology. With further refinement, this provided a means of extracting the flavour from foods and beverages during processes involving heating which often result in the loss of flavour. Much of the early research and development took place on the shop floor, backed up by collaboration with local wine makers, while later research involved working with university and other research personnel.

Because of the limited national demand for this highly specialized and expensive capital equipment, the firm needed to look towards international markets to build up the demand for this equipment. In order to give more emphasis to export markets, and to employ the specialized labour force more productively, as well as to generate investment and undertake further research and development, the division producing the spinning cone technology was hived off as a separate enterprise under the name Flavourtech (see Flavourtech webpage). The new enterprise was established in its own premises adjacent to A&G Engineering, with its own management, but with close links to A&G through some common ownership and membership of their boards of directors.

## Conclusion

This chapter has argued that Kaldor's stages model of circular and cumulative causation is an important methodological tool for understanding the development of particular firms and industries. His model does not suffer the general criticisms of stages models put forward by Gunnar Myrdal, the other significant proponent of circular and cumulative causation. By demarcating the points at which one stage moves into the next and the problems involved in this transition, Kaldor's model avoids the teleological nature of other stages theory – growth is not inevitable and its path is not pre-determined in its specific details.

The demarcation points for the stages of growth are determined by the changes in activities that firms and industries need to implement to tap larger and more geographically dispersed markets, especially with respect to capital equipment. This was evident in the wine industry cases described in this chapter. Growth within a stage involved many small-scale incremental changes, but the "step-up" into the next stage usually involved a discontinuous alteration of the firm's equipment needs, and also in other factors such as marketing and distribution methods.

We also argued that Kaldor neglected an important stage in the circular and cumulative causation process, the localized demand of regional markets which are critical at the "embryonic" stage of firm development. Other writers have argued that firms achieving success in international markets are usually firmly based in a local market, since it provides feedback from customers to improve products, and allows firms to solve problems that could prove disastrous internationally (Bartlett and Ghoshal 2000). Porter (1996: 87) comments on "the motivational benefits of local competition and the importance of sophisticated local demand for a product or service".

We believe that the regional stage of development is particularly significant in industries that rely on a key agricultural input where transport and storage problems are important; hence the choice of the wine industry as a case study. We also argue that for capital goods producers, especially those producing specialized (rather than general-purpose) equipment, local demand is an important element in the early period of their developmental trajectory. As Toner (2000: 24) points out, capital goods are particularly significant for economic growth as they provide resources for other manufacturers, and stimulate the development of technological service industries such as computer software.

The findings from the case studies highlight the significance of a firm's history at the regional stage for its future development. Although Kaldor (1972: 186) did not refer specifically to a regional stage of enterprise development, he recognized that the forces of continuous change are endogenous, and that a particular development could not be predicted "except as a result of the sequence of events in previous periods which led up to it". Young (1928: 533) had argued that a state of continuous change occurred in industry as a result of external "adventitious" elements, as well as the internal daily operations of manufacturing enterprises. Thirlwall (1987: 327) has commented that "the present and the



future cannot be understood without reference to the past”, while Setterfield (1997: 366) refers to “the heritage of the past as the only true variable in the system”.

## References

- A&G Engineering, [www.agengineering.com.au](http://www.agengineering.com.au).
- Argyrous, G. [1995] (1996) “Economic Evolution and Cumulative Causation”, in G. Argyrous and F. Stilwell (eds) *Economics as a Social Science: Readings in Political Economy*, Sydney: Pluto Press.
- Argyrous, G. and Sethi, R. (1996) “The Theory of Evolution and the Evolution of Theory: Veblen’s Methodology in Contemporary Perspective”, *Cambridge Journal of Economics*, 20: 475–95.
- Bamberry, G. (2006a) “The Significance of Linkages for Regional Manufacturers: A Case Study of the Riverina Region of New South Wales”, in P. K. Basu and G. O’Neill (eds) *Proceedings of the Third International Conference on Contemporary Business*, Leura, Bathurst: Charles Sturt University.
- (2006b) “The Influence of Technology on Regional Development: Case Studies from the Riverina Region”, *Australasian Journal of Regional Studies*, 12, 2: 173–90.
- Bartlett, C. A. and Ghoshal, S. (2000) “Going Global: Lessons from Late Movers”, *Harvard Business Review*, March–April: 132–42.
- Brokes Promise Wines, [www.brokespromise.com.au](http://www.brokespromise.com.au).
- Casella Wines, [www.casellawines.com.au](http://www.casellawines.com.au).
- De Bortoli Wines, [www.debortoli.com.au](http://www.debortoli.com.au).
- Flavourtech, [www.ft-tech.net](http://www.ft-tech.net).
- Glendonbrook Wines, [www.glendonbrook.com.au](http://www.glendonbrook.com.au).
- Kaldor, N. [1966] (1978) “Causes of the Slow Rate of Economic Growth in the United Kingdom”, in *Further Essays on Economic Theory*, New York: Holmes & Meyer.
- [1970] (1978) “The Case for Regional Policies”, in *Further Essays on Economic Theory*, New York: Holmes & Meyer.
- [1972] (1978) “The Irrelevance of Equilibrium Economics”, in *Further Essays on Economic Theory*, New York: Holmes & Meyer.
- [1981] (1989) “The Role of Increasing Returns, Technical Progress and Cumulative Causation in the Theory of International Trade and Economic Growth”, in F. Targetti and A. P. Thirlwall (eds) *Further Essays on Economic Theory and Policy: Nicholas Kaldor*, London: Duckworth.
- Kerlinger, F. N. and Lee, H. B. (2000) *Foundations of Behavioral Research* (4th edn), South Melbourne, Vic: Wadsworth/Thompson Learning.
- Martin, R. (1999) “The New ‘Geographical Turn’ in Economics: Some Critical Reflections”, *Cambridge Journal of Economics*, 23: 65–91.
- Molly Morgan Wines, [www.mollymorgan.com](http://www.mollymorgan.com).
- Myrdal, G. (1957) *Economic Theory and Underdeveloped Regions*, London: Duckworth.
- (1968) *Asian Drama*, New York: Pantheon.
- Penmara Wines, [www.penmarawines.com.au](http://www.penmarawines.com.au).
- Porter, M. (1996) “Competitive Advantage, Agglomeration Economics, and Regional Policy”, *International Regional Science Review*, 19, 1 and 2: 85–94.
- Roberts, B. (2000) “Facilitating Industry Cluster Development”, *Regional Policy and Practice*, 9, 1: 36–45.

- Scott, A. J. and Storper, M. (1992) "Industrialization and Regional Development", in M. Storper and A. J. Scott (eds) *Pathways to Industrialization and Regional Development*, London: Routledge.
- Setterfield, M. (1997) "History vs Equilibrium and the Theory of Economic Growth", *Cambridge Journal of Economics*, 21: 365–78.
- Targetti, F. (1992) *Nicholas Kaldor, the Economics and Politics of Capitalism as a Dynamic System*, Oxford: Oxford University Press.
- Thirlwall, A. (1987) *Nicholas Kaldor*, Brighton: Wheatsheaf Books.
- Ticehurst, G. W. and Veal, A. J. (2000) *Business Research Methods: A Managerial Approach*, Frenchs Forest, NSW: Longman.
- Toner, P. (1999) *Main Currents in the Theory of Circular and Cumulative Causation: The Dynamics of Growth and Development*, New York: St Martin's Press.
- (2000) "Manufacturing Industry in the Australian Economy: Its Role and Significance", *Journal of Australian Political Economy*, 45: 18–45.
- Westend Estate Wines, [www.westendestate.com.au](http://www.westendestate.com.au).
- Wickramasekera, R. and Bamberry, G. (2001) "Australian Wineries: Factors Perceived to Enhance or Inhibit Export Market Expansion", in S. J. Gray, S. L. McCaughey and W. R. Purcell (eds) *Asia Pacific Issues in International Business*, Cheltenham: Edward Elgar.
- (2003) "Exploration of Born Globals/International New Ventures: Some Evidence from the Australian Wine Industry", *Australasian Journal of Regional Studies*, 9, 2: 207–20.
- Young, A. (1928) "Increasing Returns and Economic Progress", *Economic Journal*, 38, 152: 527–42.

## 5 **Nicholas Kaldor and cumulative causation**

### Public policy implications

*Richard P. F. Holt and Steven Pressman*

#### **Introduction**

Public policy is one key area where Post Keynesian economics differs from neo-classical economics. Neoclassical economists believe in the efficacy of markets. As a result, they tend to oppose government interference in market economies. On the other hand, from a Post Keynesian perspective, markets sometimes fail to lead to optimal results. In such instances, government actions are necessary to intervene in markets and thereby improve economic performance.

The work of Nicholas Kaldor, who was concerned about the relation between theory, empirical work and public policy, helped shape the Post Keynesian approach to public policy. The theoretical work that Kaldor did with cumulative causation, and then his application of this work to specific public policy issues, shows the method and influence that Kaldor had on developing Post Keynesian public policy.

In this chapter we first explain the basic principles and the history of cumulative causation, including Kaldor's development and use of the notion of cumulative causation. We then show how Kaldor applied this theory to three areas and discuss the policy proposals that fall out of his analysis. These policy proposals, Kaldor believed, would deal with real world economic problems and would improve economic outcomes. We finally draw out some implications from Kaldor's work that may be used for future Post Keynesian analysis and for developing Post Keynesian policy recommendations that deal with contemporary economic problems.

#### **Cumulative causation – the basics**

One of the main principles of neoclassical analysis is its assumption that there is a stationary state towards which the economy is headed. This idea forms the basis of neoclassical general equilibrium theory, neoclassical game theory, and much of neoclassical thought. The notion of cumulative causation rejects equilibrium analysis and takes a dynamic approach to economic analysis. It shows how an economic system moves through historical time, how economies can

exhibit instabilities, and how each economic variable responds to changes in other variables.

In brief, cumulative processes involve a positive or negative feedback mechanism involving two or more variables. This contrasts sharply with a simple, unidirectional causal schema, where A causes changes in B, but the possibility of B having further effects on A is excluded. This simple causal sequence leads to some equilibrium point or point of rest. With unidirectional causation, changes in A lead to changes in B and things end there; the system reaches a new steady state with higher (or lower) values for the variables A and B.

With cumulative causation, the variables A and B continually impact upon each other. Changes in A will affect B, which will further affect A, again impact B, and so on. When A and B both increase, we have a virtuous cycle or positive feedback loop; when A and B both decline, we have a vicious circle or negative feedback loop; when A rises and B declines (or vice versa), we have a cyclical process. Some economic examples of this latter case include business cycles and swings in agricultural quantities and prices as described in the cobweb theorem.

## A history of cumulative causation

The principle of cumulative causation was first applied in economic analysis in the dynamic *Tableaux* of Quesnay (see Pressman 1994, 2007); but this work remained largely unknown to economists until Meek's (1963) pioneering study of physiocracy.

The idea of a cumulative economic process was revived when Wicksell ([1898] 1936) examined what happens when market and natural interest rates diverge. For Wicksell, if the interest rate charged by banks (the market rate) was less than the rate of return on investment (the natural rate), firms would continue to borrow money and expand, and economies would continue to grow. On the other hand, if the market rate exceeded the natural rate, investment would come to a halt, and the economy would stagnate, further reducing investment and GDP. Wicksell influenced both Keynes and the Austrians. He provided Keynes with a possible explanation for why economies can experience prolonged periods of growth, or suffer from depression or recession, and he provided the Austrians with a foundation for their theory of the trade cycle.

It was Myrdal, however, who first described the principle of cumulative causation in detail, named it, and recognized its importance for policy issues. Myrdal (1933) argued that Wicksell's cumulative process was inconsistent with equilibrium analysis, and that Wicksell ignored the importance of both uncertainty and time in cumulative economic processes. He then used the idea of cumulative causation for policy analysis.

Myrdal (1957) employed the principle to help explain persistent underdevelopment. In poor nations, the young, the talented, the ambitious and the educated will leave for more developed areas where they can earn more and enhance their skills. This brain drain leaves a small tax base to generate funds for needed

services and regional development. Moreover, domestic savings will tend to be invested abroad, where the returns are likely to be higher. As the economy stagnates, while other economies thrive, incentives grow for people to move to developed countries. In contrast to the predictions of the Samuelson–Stolper theorem, where incomes throughout the world become more equal due to the movement of resources, for Myrdal poverty persists in underdeveloped areas and the gap between rich and poor nations grows over time – unless something is done to help underdeveloped countries and reverse the cumulative process of decline.

Myrdal (1944) also applied the idea of cumulative causation to the problem of black poverty in the US. He saw a vicious cycle facing black Americans and leading to their high and intractable poverty rates. In brief, discrimination leads to worse economic, health and education outcomes (all mutually reinforcing) for blacks; this, in turn, reinforced the prejudices that led to the discrimination against blacks, leading to further declines in the living standard of blacks relative to whites. As Myrdal (1944: 387) put it so succinctly, “*Discrimination breeds discrimination*”. Again, his solution was to have the government provide assistance to black Americans (such as better health and education) in order to break the vicious cycle they faced.

At the heart of this application of cumulative causation is the idea of increasing returns – that productivity or efficiency improves as production increases. This idea was developed by Young (1928), who credits the idea to Adam Smith’s ([1776] 1937) pin factory, but the seed of this notion really goes back to Quesnay and the Physiocrats. Quesnay argued for favorable treatment of France’s agricultural sector, believing that greater demand for agricultural goods would spur efficiency gains and improvements in agricultural methods (Pressman 1994). Increasing returns to scale in agriculture means that as agricultural output increases, productivity rises in the agricultural sector, leading to greater food production. This means that real incomes will grow and spending will grow. As a result, all economic sectors flourish. Economic growth then generates continued productivity gains and further improvements in all the variables. In contrast, when output falls in the agricultural sector, productivity growth falls there, and incomes decline throughout the economy. This reduces spending and generates continued economic decline. Following Smith, for Young and Kaldor, manufacturing, rather than agriculture, exhibited increasing returns and was the engine for economic growth; but the process was essentially the same – a growing economic sector, experiencing increasing returns, generates a cumulative process of growth and improved living standards.

### **Kaldor and his theory of cumulative causation**

Kaldor was likely unaware of the analysis of Quesnay and the Physiocrats; his work contains no references to Quesnay’s dynamic *Tableaux* or to the main ideas of the Physiocrats. But Kaldor was certainly aware of the work of Young, Wicksell and Myrdal.

Kaldor was a student at the London School of Economics from 1927 to 1929 when Young was teaching there. At this time Young was working on his famous increasing returns paper, which he delivered as President of the British Association at the University of Glasgow in September 1928. Young usually lectured on whatever topic he was currently working on, so it is pretty certain that Kaldor was exposed to the notion of increasing returns through Young (Blitch 1995: 164–9, 178f.). Kaldor regularly references the work of Young on increasing returns, and he frequently cites increasing returns as a main factor leading to cumulative causation. The only thing Young missed, according to Kaldor (1972: 1249), is the Keynesian income effects stemming from increasing returns.

Kaldor was also aware of the work of Wicksell. Kaldor (1938, 1939, 1940, 1942) refers to Wicksell continuously in his early papers on trade and on interest rates. Hayek, too, was teaching at the LSE when Kaldor was there, and Kaldor (1939) attacked Austrian theory in one of his early papers on the trade cycle and Hayek in another (Kaldor 1942).

Finally, Kaldor was familiar with the work of Myrdal. Hicks introduced Kaldor to Myrdal's (1933) *Monetary Equilibrium* (Thirlwall 1987: 25). Kaldor also worked for Myrdal in the 1940s as Director of the Research and Planning Division of the Economic Commission for Europe, part of the UN, created to administer Marshall Aid and assist in the reconstruction of Europe after the Second World War. And he quotes Myrdal (1957) in his paper (Kaldor 1972) opposing equilibrium analysis.

With these influences on his thinking, it is not surprising that Kaldor (1972, 1985) would have serious concerns about equilibrium analysis. His first published article (Kaldor 1934) contained an account of the cobweb theorem, clearly showing dissatisfaction with equilibrium theorizing, and argued that equilibrium is undetermined in many real world instances. In some of his earliest papers, Kaldor (1939, 1940) used the idea of cumulative causation to show how economies could go through a series of business cycles, while at the same time the economy moves along on either an upward or downward trajectory. Later, Kaldor (1985: 24f.) would emphasize that quantity changes are more important than price changes, and that this caused cyclical behaviour in commodity markets and at the macroeconomic level.

As Thirlwall (1987: 319ff.) points out, Kaldor had three main objections to equilibrium theory. The first objection was methodological – the theory was tautological and could not be tested or proven wrong. The second objection was that it relied too much on substitution effects and the allocative function of markets. Because of its assumption of full employment and efficient resource allocation, any change must have opportunity costs. Allowing for the possibility of underutilized resources, or replacing substitution effects with income effects, means that there can no longer be a neat equilibrium solution. In the real world, where increasing returns operate, capital and labor become complementary, the manufacturing and the service sectors expand together, as the expansion of one sector generates demand for the goods of other sectors. Forces for change are endogenous in the system, and there is a cumulative process of change (Thirlwall 1987: 322). Finally, in the presence of increasing

returns there can be *no* movement toward equilibrium and the idea of optimal allocation of resources is meaningless since the position of a production possibility curve depends on the allocation of resources. Simply put, as more resources are devoted to capital goods, production possibilities expand. Increasing returns also means that wages and employment will be positively related and mutually reinforcing. Increased wages lead to increased demand, increased employment, and higher wages.

For Kaldor (1985), cumulative causation, rather than equilibrium analysis, means that economies follow a growth process with no mechanism to establish full employment and no equilibrium growth rate that would tend to establish full employment. Jettisoning equilibrium, economic analysis must understand how actual economies work and what causes them to change over time.

This approach has important policy implications. Economics without equilibrium is not an economics to be left to the invisible hand, as Adam Smith advocated. Rather, it follows the lead of Quesnay and Myrdal, who argued that government action is necessary to promote economic growth. It requires economic policies to promote a virtuous circle, improve human well-being, and enhance overall economic performance. It is an economics that uses policy measures to improve both efficiency and equality.

## **Kaldor's approach to policy issues**

Kaldor's work in critiquing equilibrium theory and in developing the theory of cumulative causation has a number of specific policy implications. Many of these Kaldor himself drew out during the course of his productive career. These policy implications are quite different from those that follow from neoclassical economic theory.

Kaldor is probably best known for three main policy proposals, which we focus on below. These policies all follow from an economic analysis focusing on economies moving through time and employing the principle of cumulative causation – an industrial policy to promote economic growth in England, an expenditure tax to promote greater equality as well as long-run growth, and an incomes policy (rather than monetarism) as a means of controlling inflation.

## **An industrial policy**

Kaldor's analysis of economic growth through the use of cumulative causation, and his critique of neoclassical economics, offers a different approach to the problem of economic growth. Kaldor, like other Post Keynesians, focuses more on demand or income effects than on supply-side substitution effects when analyzing economic growth. He also incorporates feedback effects, or cumulative causation, into his analysis. On this view, an increase in demand leads to economic growth, more demand, and more growth.

The first argument of the importance of demand for productivity growth comes from Adam Smith ([1776] 1937). For Smith, the greater the extent of the



market, the greater the amount of sales and the greater the growth in productivity through the division of labour. This idea lied dormant in economics until the early twentieth century when Allyn Young (1928) argued that many industries operate under conditions of increasing returns to scale. As noted earlier, this is an idea that Kaldor developed and incorporated into his theory of growth. As more and more is produced, costs fall because fewer resources are used to produce each good. Increasing returns thus means that productivity grows as output expands. In general, during times of slow economic growth, productivity will not grow by much. Following Smith, Young believed that larger markets and greater sales would lead to the greater use of automation and thus to increasing returns in large parts of the economy. Hence, it is the growth of demand that determines productivity growth.

Kaldor further developed the theory that demand determines productivity growth by suggesting that productivity growth stems from recognizing that economic sectors differ in their efficiency or productivity. When demand shifts to goods produced by more productive economic sectors, average productivity levels will increase. Kaldor (1966; 1967) thus stressed the composition of demand as a factor affecting productivity growth. Like Smith and Young, he looked at the manufacturing sector rather than at the agricultural sector as the engine for productivity growth, but in essence his argument was the same as that of Quesnay – some economic sectors are inherently more productive than other economic sectors, and government should help more productive economic sectors to grow.

For Kaldor, cumulative causation was an integral part of the growth process. Those nations that developed their manufacturing sector embark on a virtuous cycle of productivity and income growth; in contrast, those nations specializing in agriculture or services will experience stagnating productivity and incomes, and a vicious cycle of decline.

Thirlwall (1983) has identified several propositions at the core of Kaldor's theory. First, and most important, the growth of the manufacturing sector determines productivity growth in the manufacturing sector. Second, because of spin-offs from the manufacturing sector, the greater the growth of the manufacturing sector, the greater the growth of productivity outside the manufacturing sector. Finally, according to Kaldor, the growth of the manufacturing sector is not supply-constrained but demand-constrained. Manufacturing production is determined primarily by overall demand in the economy and by (relative) demand for manufactured goods. Growing demand usually leads to even faster demand for manufactured goods due to the high elasticity of demand for manufactured goods (Eatwell 1982). The greater demand for manufactured goods boosts productivity in the manufacturing sector (due to increasing returns) and productivity growth for the whole economy (due to spillover effects and the nature of weighted averages).

The argument for this was both theoretical and empirical. At the theoretical level, Kaldor relied on the existence of increasing returns in manufacturing production and the large scope for technological change in manufacturing industries. In addition, demand for manufactured goods (unlike agricultural goods)

keeps rising as income increases. At the empirical level, Kaldor identified a number of stylized facts or real world regularities to support his position about the importance of the manufacturing sector. First, he found a high correlation between economic growth and the growth of manufacturing output for 12 industrial countries during the 1950s and 1960s. He argued that aggregate growth depended upon manufacturing growth (rather than the other way around), and that this could be explained by increasing returns in manufacturing. Second, Kaldor found a high correlation between productivity growth in the manufacturing sector and the growth of manufacturing output. He argued that economies of scale means that productivity growth accelerates and costs fall in manufacturing when demand for manufactured goods rises.

One policy implication of this analysis is that economic policies must support the British manufacturing sector in order to spur growth in Britain. Governments could provide tax breaks to manufacturing firms, especially in depressed areas (Kaldor 1970b; this policy is a forerunner of enterprise zones), or provide regulatory relief to them. Britain could also protect and support domestic manufacturing firms against foreign competition. Here Kaldor recommended devaluation of the pound, import controls, and keeping Britain out of the Common Market.

These last policies also contributed to export-led growth for Britain. Export-led growth has Keynesian multiplier effects and also, due to increasing returns, increases labor productivity. This will lower costs, and given Kaldor's (1956) view that firms engage in markup pricing, would lower prices for UK goods. This, in turn, will increase exports and our virtuous circle continues to proceed in this manner.

Similar to this, many US economists advocated industrial policy in the 1980s (also see Magaziner and Reich 1982; Reich 1983). They argued that within the manufacturing sector, a certain set of manufacturing industries (such as automobiles, consumer electronics, and computers) would create more value for each worker than other industries. They then argued for an economic policy that would actively favour more productive economic sectors in order to increase incomes, aggregate demand and economic growth.

One unique, but untried, policy proposal that followed from Kaldor's analysis of cumulative causation, productivity growth, and the manufacturing sector was the selective employment tax. Kaldor (1960–80, Vol. 7: 200–9; 1966: ch. 7) proposed that firms in the service sector should be taxed based on the number of workers they employed. This would discourage employment outside manufacturing and encourage employment in manufacturing industries experiencing increasing returns to scale. This too would stimulate a cumulative process of income growth and demand growth

All of these policies should be contrasted with the neoclassical approach to economic growth, which focuses mainly on the supply side of the economy and seeks to generate the proper economic incentives to spur growth, after which the government just stands on the sidelines. As a result, neoclassical growth theory advocates tax cuts for the wealthy to generate more investment, entrepreneurship, and hard work. Likewise, the neoclassical approach calls for cuts in social benefits so that

people will have incentives to work hard and produce more rather than merely collecting these benefits. Overall, the neoclassical approach opposes having the government make economic decisions by supporting one sector over another and opposes having the government maintain a permanent role in the economy by continually supporting economic growth through activist policy measures.

### **An expenditure tax**

Cumulative causation arguments also support Kaldor's famous proposal to replace the income tax with an expenditure tax. The idea of taxing expenditures rather than income has received lukewarm support among Post Keynesians, who tend to see the plan as a way to reward savings and discourage the spending that drives economic growth. Kaldor himself saw the expenditure tax as a way to raise savings and investment by focusing on increasing savings, and he emphasized the gains from encouraging savings in his book on the expenditure tax. Kaldor (1955: 53) approved the idea of taxing people based on what they took out of the system rather than what they put in, and Kaldor (1955: 84) talked about the expenditure tax as something that would encourage people to postpone consumption.

But this response falls prey to seeing the expenditure tax as having one and only one effect – converting consumer spending to saving as a result of changing economic incentives. In this light, the main impact of the expenditure tax would be to lower effective demand and reduce economic growth, making it an anti-Keynesian policy.

However, this simple view of the expenditure tax ignores any potential income effects. In practice it is easy to design an expenditure tax that is as progressive as one wishes a tax system to be. All that is necessary is that we manipulate the marginal tax rates on spending to yield the desired distributional consequences. For example, the poor can be given substantial tax rebates, as exists in the current US tax system via the earned income tax credit. Or the tax rate on the first several thousand dollars of spending can be made negative, as in negative income tax plans (Friedman 1962: 177–95; Tobin *et al.* 1967). And the very wealthy can be assessed at extremely high rates so that the government obtains the same amount, or even greater revenue, from them.

Kaldor (1955: 241) did recognize that just to keep the distribution of the tax burden the same under an expenditure tax, it would require marginal tax rates on consumption exceeding 100 per cent, perhaps going as high as 300 per cent. But he also supported making the tax system *more* progressive via the expenditure tax by raising the top tax rates on expenditures. Kaldor (1955: 15) argued that an expenditure tax could lead to a more egalitarian society, a point he stressed throughout his book on the expenditure tax. As Turner (1993: 45) notes, Kaldor's "argument for an expenditure tax was an equity argument more than anything else".

Further, there may be other growth-enhancing aspects of an expenditure tax that are Post Keynesian in nature but not mentioned by Kaldor. If the wealthy tend to spend more of their income on services, and to spend relatively more of their income on unique and expensive goods that are not mass produced (see

Frank 1999), then a movement toward the production of mass-produced manufacturing goods would lead to a cumulative process of productivity growth and economic growth.

Kaldor did *not* explicitly make any of these arguments in his expenditure tax book of 1955. Nor did he make them when he was running around the world advocating an expenditure tax to developing countries. Nonetheless, this defence of the expenditure tax is consistent with his thought and with Post Keynesian thought. Kaldor (1956) argued that the rate of spending out of wages was greater than the rate of spending out of profits. Thus redistributing disposable income from the wealthy (who mainly receive profit income) to others should increase the overall MPC. This increase in spending would have effects on both economic growth and business investment. As Fazzari and Mott (1986–7) argue, higher levels of consumption should increase capacity utilization and investment, and they present good empirical evidence that this cumulative process occurs in the real world.

In addition, Keynes ([1936] 1973: 95) famously noted that “If fiscal policy is used as a deliberate instrument for the more equal distribution of incomes, its effect on increasing the propensity to consume is, of course, all the greater.” For this reason he advocated high tax rates on unearned income, capital gains and inheritances, all forms of income that are received disproportionately by the wealthy. Kaldor, of course, would have known about Keynes’s argument, having been part of the Cambridge Circus. Moreover, this view has considerable empirical support. Both Pressman (1997) and Brown (2004) have both shown that income distribution does affect aggregate consumption and economic growth as the Post Keynesians have contended.

Taking into account both the substitution effects of discouraging consumption and encouraging savings, and the income effects of making the tax system more progressive, it is not clear that the expenditure tax would adversely affect economic growth. While an expenditure tax would generate incentives to save, it would also redistribute the tax burden from the middle class and the poor to the wealthy. By giving more after-tax income to those who are likely to spend it, an expenditure tax would encourage consumption, investment, and a virtuous cycle of economic growth.

## **An incomes policy**

Kaldor relied on the principle of cumulative causation to argue against monetarism during the 1970s. For monetarists, inflation results when central banks create too much money. Excessive money in circulation increases the demand for goods and services, and with limited resources to produce more goods (i.e., assuming full employment), the main result can only be higher prices.

One problem with the monetarist point of view, according to Kaldor (1970a), was that it failed to explain how additional money gets into circulation. Friedman’s simple story of central bankers dropping money from helicopters is neither true nor realistic. More importantly, monetarism ignores the fact that all

macroeconomic variables are highly correlated with each other and that changes in one variable will affect other variables. In essence, Friedman assumed unidirectional causation between changes in the supply of money and changes in the price level.

Kaldor (1976) did not deny a causal linkage going from money creation to higher prices for primary goods or commodities, which would then get passed along to consumers by firms using these materials in manufacturing and pricing based on cost considerations. However, he emphasized that while money affects prices, there was another mechanism going from changes in the price level to changes in the money supply. The problem with Friedman is that he assumed money demand was stable and so his entire position rested on a causal chain going from changes in the money supply to changes in output, but not the other way around (Kaldor 1982: 25–6).

The next step was to show and explain how changes in output lead to changes in the money supply. Kaldor (1970a, 1982) presented evidence of changing money demand when output changes, and argued that the supply of money was endogenous because banks accommodated greater demand for money. When prices increase, people and firms need to borrow more money. People need to borrow money because workers spend almost all their income. When the price of goods rises, to maintain their accustomed standard of living, people must borrow money in order to afford the higher prices. Firms too must borrow to pay for the higher costs of factors of production. Banks, wanting to keep customers happy and wanting to earn money from additional lending, do everything possible to meet this credit demand. Since increased borrowing leads to greater money creation, we have the makings of a cumulative process whereby changes in the money supply lead to changes in prices, which in turn result in greater demand for borrowed money in order to pay for the higher priced goods.

From here Kaldor went on to critique the monetarist solution of controlling the money supply as a means of controlling inflation. He noted that the cumulative process could be slowed down at any point. The question was where and how to do so. He felt that the monetarist solution of reducing money growth would lead to unemployment and that this was a high cost to pay for mitigating a rather minor problem, a general rise in prices. For Kaldor (1976: 708f.), the root cause of inflation was not excessive money creation, but a cumulative process in which various groups sought to obtain more output than what was actually produced. If both workers and firms sought to increase their share of the economic pie at the same time, the only result would be a wage-price spiral. Workers would seek more of the pie by demanding higher wages and firms would demand more by raising their prices. Together, they would produce the cumulative process known as the wage-price spiral, with additional money created merely as a byproduct.

Given this analysis, the appropriate solution would be to focus on stopping the wage-price spiral. For this reason Kaldor (1982: 61–5; 1985: 39f.) advocated an incomes policy to deal with the inflationary pressures. If workers agreed to accept lower pay (or pay increases in line with labor productivity growth), and

firms agreed to keep prices constant in the face of constant costs, relative incomes would stay the same and inflation would be brought under control. This, in turn, would lower the demand for loans on the part of firms (to pay higher wages) and on the part of consumers (to pay for higher prices), thereby reducing the money supply.

Kaldor's other policy analyses provide further arguments for an incomes policy. Lower inflation would help increase exports, thereby generating productivity gains that would further reduce inflationary pressures. A steeply progressive expenditure tax would support these efforts since the after-tax gains from higher wages and profits would be worth much less when spent on goods and services. Seen in this light, an expenditure tax begins to look a lot like a tax-based incomes policy (see Wallich and Weintraub 1971; Neale 1986). Under this plan, those with wealth and economic power are penalized for exerting this power in antisocial ways that push up prices for everyone. This is especially true when expenditure tax rates are extremely high for the well-to-do.

## Learning from Kaldor

It is unfortunate that Post Keynesians have not devoted more time and effort to the important issue of productivity growth as Kaldor did. It is also unfortunate that they have not pushed their demand-side analysis of productivity growth and economic growth. These topics are ripe for Post Keynesian analysis and for a Post Keynesian set of economic policies. Kaldor's unique perspective helps Post Keynesians not only to distinguish themselves from neoclassical economic thought but also from other heterodox approaches.

The fact that the data seems to support Kaldor's approach is another reason to pursue this analysis further. As we have seen, productivity growth has slowed in most of the developed world just at the time when economic growth rates have slowed and unemployment rates have increased. Moreover, the fact that productivity growth in the US seems to have turned around in the late 1990s, just as the US economy was growing and unemployment was falling to near full employment levels, adds further empirical support to Kaldor's position.

There still remains a great deal of work to do on this topic. First, careful empirical analysis, building on the work of Kaldor, needs to show that high unemployment and low growth are critical factors in the productivity slowdown of the late twentieth century. An important and related issue is whether *stable* unemployment rates and *stable* growth rates also contribute to productivity growth. To the extent that such stability reduces uncertainty and increases business investment and consumer spending, it should spur productivity growth.

Second, it is still not clear which are the more productive economic sectors, whether these sectors change over time, and what causes such changes. Most Post Keynesians, following Kaldor, have focused on the manufacturing sector as the engine of growth and the source of productivity improvements. But advocates of a "new economy" focus on computer technology, bio-technology, and especially the services related to this technology, as the key areas of the world

economy. If we are going to assist productive sectors through economic policy, and if we are going to use economic policy to generate a cumulative process of economic growth, it is important to correctly identify these more productive economic sectors.

Assuming we can identify the appropriate sectors, a key policy issue is how to structure economic policy so that these key sectors are favoured over other sectors. This may include tax breaks for these sectors, government purchases of the goods produced by these sectors, loan guarantees, or other forms of assistance. Political backlash, stemming from the inequity of favouring some economic sectors over others, will need to be addressed. So too will the fact that government policies to advance any economic transformation will create losers. These losers are likely to become disgruntled voters who may try to overturn any economic policy that would ultimately lead to the greater public good. But they are also human beings whose needs during the transition must be considered and must be dealt with in a fair and humane way. Thus, any analysis of the political economy of productivity growth and sectoral change must make sure that the transition of labor and capital between different sectors is as smooth as possible.

## Summary and conclusion

Kaldor's unique contribution to the history of economic thought and economic policymaking was to take existing arguments about the importance of cumulative economic processes and apply this framework to key policy issues. As a result of viewing the economy as a dynamic process, he developed several policy proposals to help put domestic economies on a virtuous trajectory of economic growth. These policy implications stand in sharp contrast to the standard neo-classical policy approach, which is essentially a *laissez-faire* approach. It is time for Post Keynesians to follow Kaldor's lead and begin to address these key issues of our time.

## References

- Blicht, C. (1995) *Allyn Young: The Peripatetic Economist*, New York: St Martin's Press.
- Brown, C. (2004) "Does Income Distribution Matter for Effective Demand? Evidence from the United States", *Review of Political Economy*, 16: 291–307.
- Eatwell, J. (1982) *Whatever Happened to Britain?*, New York: Oxford University Press.
- Fazzari, S. and Mott, T. (1986–7) "The Investment Theories of Kalecki and Keynes: An Empirical Study of Firm Data, 1970–1982", *Journal of Post Keynesian Economics*, 9: 171–87.
- Frank, R. (1999) *Luxury Fever*, New York: Free Press.
- Friedman, M. (1962) *Capitalism and Freedom*, Chicago, IL: University of Chicago Press.
- Kaldor, N. (1934) "A Classificatory Note on the Determinateness of Static Equilibrium", *Review of Economic Studies*, 1: 122–36.
- (1938) "Stability and Full Employment", *Economic Journal*; reprinted in *N. Kaldor – Essays on Economic Stability and Growth* (1960), Glencoe, IL: Free Press.



- (1939) “Capital Intensity and the Trade Cycle”, *Economica*; reprinted in *N. Kaldor – Essays on Economic Stability and Growth* (1960), Glencoe, IL: Free Press.
- (1940) “A Model of the Trade Cycle”, *Economic Journal*; reprinted in *N. Kaldor – Essays on Economic Stability and Growth* (1960), Glencoe, IL: Free Press.
- (1942) “Professor Hayek and the Concertina-effect”, *Economica*; reprinted in *N. Kaldor – Essays on Economic Stability and Growth* (1960), Glencoe, IL: Free Press.
- (1955) *An Expenditure Tax*, London: Unwin University Books.
- (1956) “Alternative Theories of Distribution”, *Review of Economic Studies*, 23: 83–100.
- (1960–80) *Collected Economic Essays, 9 Vols*, London: Holmes & Meier.
- (1966) *Causes of the Slow Rate of Economic Growth in the United Kingdom*, Cambridge: Cambridge University Press.
- (1967) *Strategic Factors in Economic Development*, Ithaca, NY: New York State School of Industrial Relations.
- (1970a) “The New Monetarism”, *Lloyd’s Bank Review*, 97: 1–17.
- (1970b) “The Case for Regional Policies”, *Scottish Journal of Political Economy*, 17: 337–48.
- (1972) “The Irrelevance of Equilibrium Economics”, *Economic Journal*, 82: 1237–55.
- (1976) “Inflation and Recession in the World Economy”, *Economic Journal*, 86: 703–14.
- (1982) *The Scourge of Monetarism*, Oxford: Oxford University Press.
- (1985) *Economics without Equilibrium*, Armonk, NY: M. E. Sharpe.
- Keynes, J. M. ([1936] 1973) *The Collected Writings of John Maynard Keynes*, Vol. VII, *The General Theory of Employment, Interest, and Money*, London: Macmillan.
- Magaziner, I. and Reich, R. (1982) *Minding America’s Business*, New York: Harcourt Brace Jovanovich.
- Meek, R. (1963) *The Economics of Physiocracy: Essays and Translations*, Cambridge, MA: Harvard University Press.
- Myrdal, G. (1933) *Monetary Equilibrium*, London: W. Hodge & Company.
- (1944) *An American Dilemma*, New York: Harper & Brothers.
- (1957) *Economic Theory and Underdeveloped Regions*, London: Duckworth.
- Neale, W. (1986) “Tax-based Incomes Policies: A Commentary for the Future”, *Journal of Economic Issues*, 10: 969–87.
- Pressman, S. (1994) *Quesnay’s Tableau Économique: A Critique and Reconstruction*, Fairfield, NJ: Augustus Kelly.
- (1997) “Consumption, Income Distribution and Taxation: Keynes’s Fiscal Policy”, *Journal of Income Distribution*, 7: 29–44.
- (2007) “Francois Quesnay: The First Post Keynesian?”, in M. Forstater, G. Mongiovi and S. Pressman (eds) *Post Keynesian Macroeconomics: Essays in Honour of Ingrid Rima*, London and New York: Routledge.
- Reich, R. (1983) *The Next American Frontier*, New York: Times Books.
- Smith, A. ([1776] 1937) *The Wealth of Nations*, New York: Modern Library.
- Thirlwall, A. (1983) “A Plain Man’s Guide to Kaldor’s Growth Laws”, *Journal of Post Keynesian Economics*, 5: 345–58.
- (1987) *Nicholas Kaldor*, New York: New York University Press.
- Tobin, J., Pechman, J. and Mieszkowski, P. (1967) “Is a Negative Income Tax Practical?”, *Yale Law Journal*, 77: 1–27.
- Turner, M. (1993) *Nicholas Kaldor and the Real World*, Armonk, NY: M. E. Sharpe.

- Wallich, H. and Weintraub, S. (1971) "A Tax-based Incomes Policy", *Journal of Economic Issues*, 5: 1–19.
- Wicksell, K. ([1898] 1936) *Interest and Prices*, London: Macmillan.
- Young, A. (1928) "Increasing Returns and Economic Progress", *Economic Journal*, 38: 527–42.

## 6 The principle of circular and cumulative causation

### Myrdal, Kaldor and contemporary heterodox political economy<sup>1</sup>

*Phillip Anthony O'Hara*

#### Introduction

Circular and cumulative causation (CCC) has been a critical principle of political economy for over a hundred years. While the roots of the concept go back further (see Humphrey 1990; O'Hara 2000), Thorstein Veblen (1857–1929) used the concept in his examination of the evolution of institutions. Gunnar Myrdal (1898–1987) scrutinized the conditions of African Americans and Asian under-development through the lens of CCC; influenced as he was by Knut Wicksell (1851–1926) (Myrdal 1939). Nicholas Kaldor (1908–86) applied CCC to the role of manufacturing in capitalist growth; he was influenced by Adam Smith (1723–90) and Allyn Young (1876–1929). Numerous other scholars have used the notion of CCC, often in different ways. There are linkages between Veblen, Myrdal and Kaldor. For instance, Veblen influenced Allyn Young who in turn taught Kaldor; Myrdal got the concept from Knut Wicksell and worked with Kaldor at the United Nations (Economic Commission for Europe); and Kaldor (1970: 142) got the term from Myrdal.

The first main section studies the similarities and differences between the Myrdalian and Kaldorian CCC frameworks. The second section develops a general system model of integration between the two traditions. There is a large measure of continuity between the two CCC approaches; they complement each other. Myrdalian CCC concentrates on the social provisioning aspect of development, while Kaldorian CCC centres on demand–supply relationships linked to the manufacturing sector. Linking both CCC approaches in an integrative model enhances our understanding of development and growth dynamics, and contributes to the development of institutional-evolutionary political economy. The third section illustrates a contemporary application of CCC through an investigation of the dynamic forces expanding the scope, network interaction and conceptual-empirical edifice of heterodox political economy.

## Comparison of Myrdal and Kaldor on CCC

Myrdalian and Kaldorian CCC traditions have significant commonalities as well as important differences. They have three main things in common. The first is the notion of circular causation, where the variables are interrelated, and the general manner of interaction between variables is complex and manifold. Circular causation is a multi-causal approach where the core variables and their linkages are delineated. CCC eschews single factor theories (O'Hara 2007a). Both Myrdalian and Kaldorian CCC examine circular relationships, where the interdependencies between factors are relatively strong, and where variables interlink in the determination of major processes.

The second similarity is cumulative causation, where the variables tend to operate as positive feedback processes, magnifying and multiplying the combined impact of the interactions through historical time. The coefficients of interaction between variables will play some role here, as will the extent of any negative feedback (drawback) effects working in the opposite direction. These circular interactions are crucial to Myrdalian and Kaldorian empirical studies of money, growth, demand, development and ethnicity. Both forms of CCC examine cumulative dynamics, where the feedback within and between variables tends to often have a multiplier or amplified impact on the overall outcomes.

The third similarity relates to traverse, path-dependence and hysteresis that move the system through time in a typically non-equilibrium fashion (Setterfield 1997). Both approaches to CCC recognize the importance of history and time, as well as space and geography, since changes to the social and political economy condition the path of evolution and transformation; and there are regional differences to growth and development as well. The acquisition of knowledge, technical skills and economies of scale affect the path of growth and development in complex and multifarious ways. Both theories explain real world processes that impact upon nations and regions, and which help explain differences in the outcomes between regions and areas.

The fourth similarity is that cumulative processes often have endogenous contradictions embedded in their dynamics. This aspect has been underemphasized in the literature, yet it is very important since it means that cumulative changes may sow the seeds of their own demise. When David Gordon (1991), for instance, criticized Kaldor's theory for having too much cumulation and not enough contradiction, he was cognizant of the problem but underplayed the degree that Kaldor himself recognized the problem (e.g. see Kaldor 1966). Setterfield (2001) has set the record straight for Kaldor, since, for instance, regimes of accumulation often have norms and mores that become locked-in, even when industrial change is required (see also Argyrous 2001; Toner 2001). For Myrdal, on the other hand, the contradictions are more obvious, since cumulation occurs more specifically in tandem with uneven development; and counteracting forces can often be strong (though themselves cumulative, perhaps in a different direction).

These are strong similarities; core ones. Indeed, they are the foundation for linking the traditions. However, the differences are also important, since they

allow the traditions to examine marginally different (but complementary) problems. There are three main differences between the models; differences of emphasis rather than of quality. The first is that Myrdalian CCC concentrates on the *social economy* and development through interdisciplinary analysis; whereas Kaldorian CCC centres on more technical demand–supply issues linked to economies of scale and growth. Although Myrdal started out applying CCC to money and macroeconomics (Myrdal 1939), his most famous two-volume application was to the underprivileged situation of African Americans in the US (Myrdal 1944), along with his three-volume work on Asian underdevelopment (Myrdal 1968). Myrdal influenced others to apply the theory to issues such as the provision of public and social services in rural and remote areas (Fagence 1980), the socio-political crisis in Poland in the 1980s (Tarkowski 1988), and uneven development at the regional level (Higgins and Savoie 1995). Myrdal's holistic vision is consistent with an interdisciplinary method for the social sciences, broadening the field of inquiry to social, political and economic relationships (see Hawley 1979).

Kaldor's CCC was a narrower economic approach to linking demand with supply through interdependencies with investment spending, productivity and world income. He placed more emphasis than Myrdal on the growth impact of CCC processes in domestic, regional and world economies. Kaldor (1972, 1975, 1980) recognized the importance of history and time, especially investment demand being embodied in scaled economies and regimes of accumulation. He stimulated other economists to apply his analysis to issues of industrial maturation and demise in the UK (Eatwell 1982), the balance of payments constraint (McCombie and Thirlwall 1994) and regimes of accumulation (Pini 1995). Kaldor's vision is narrower than Myrdal's; yet it still recognizes the importance of multi-causal processes and long-term change.

Second, the Myrdalian system is more values-oriented, concerned with the role of ideology, assumptions, social norms and mores; whereas the Kaldorian system is seemingly more objectively founded on empirical evidence (Berger 2008). Myrdal thus emphasizes the normative elements of inquiry, recognizing the role of human relationships and psychological preconceptions in the grounding of economic processes. Kaldor, on the other hand, to some degree takes for granted the cultural fabric and psychological foundations of human behaviour; concentrating on the more obvious productive, sectoral and organizational linkages within the economy.

Third, the Myrdalian system concentrates more on the uneven process of development, especially vis-à-vis minority groups and underdeveloped nations (e.g. African Americans; Asian social economies). The Kaldorian model, on the other hand, is a forward-looking view of CCC as the driving force of capitalism, for those regions that are pushing ahead of the pack. Therefore, Myrdal centres on the groups and nations or areas that are less wealthy; whereas Kaldor's core concern is the differential in terms of the forward (and contradictory) motion of the dominant areas.

It is also true that the above differences are overlapped somewhat and that Myrdal and Kaldor had much in common even at these levels, as Toner (1999: 110–12, 115–16, 159) emphasizes. For instance, Myrdal emphasized economies of scale, but with an institutional flavour; while Kaldor at times looked to an institutional and political explanation. Institutional aspects of Kaldor's theory are very important, even if seldom recognized. Hodgson (1989) argued that the highly mechanical nature of manufacturing may stimulate economies of codified knowledge; and that manufacturing can more easily aggregate economies in compact spatial centres, resulting in greater diffusion and dissemination of knowledge. It is also possible that institutional and historical factors associated with flexible norms and routines, as well as changes to the socio-economic environment, may stimulate a productivity edge for some nations and areas. Indeed, Kaldor's (1966: 110–12) emphasis on the manufacturing sector often linked to transport, utilities and communications externalities.

The similarities between Myrdal and Kaldor help us to recognize their common method and practices; while the differences (which concern degrees and core concerns) enabled them to concentrate on somewhat different levels of analysis. Myrdalian and Kaldorian differences complement each other when it comes to CCC. This is so because Myrdal concentrates on the social economics of CCC while Kaldor centres on the supply–demand dynamics of CC. A degree of specialization was thus possible, where Myrdal and associates could concentrate on socioeconomic development and inequality, while Kaldor and colleagues centred on laws of manufacturing and supply–demand interactions. Both types are well developed, and from this knowledge base we are able to detail their manner of *interaction* (see O'Hara 2007b).

Due to linkages and innovations introduced into Myrdalian and Kaldorian CCC it is useful to merge the two traditions. This merger will widen the sphere of knowledge and application within contemporary political economy. The following section illustrates this fusion into a general Myrdalian–Kaldorian system of CCC.

## General system integration of Myrdalian–Kaldorian CCC

Here we develop a dual model of interaction between the social economy and demand–supply conditions for growth and development. This model develops with the following integral conditions and assumptions for linking Myrdalian and Kaldorian CCC:

- 1 *Values and culture*: The starting point of CCC is the analysis of the role of culture in the socio-economic process. We need to recognize the significance of ideology, paradigms, human relationships and various norms and mores. We also recognize that the real world does exist, and that the critical task is to situate the causal linkages between these elements of the social economy.
- 2 *Stylized facts*: It follows logically that if we first understand the causal linkages between economic agents' valuation of the facts and their interactive relationships, then we can go forward and situate the more technical ele-

ments of the stylized facts (Skott 1999). Economic theory should be realistic in trying to base policies and practices on empirical regularities and stylized facts. Institutions, industries and trends are the foundation for a pragmatic political economy of capitalism and its alternatives. Results considered now may change in the future as new processes and transformations come into play.

- 3 *Multi-factor approach*: Single factor theories should be eschewed in favour of broader approaches that recognize the importance of the social, political and economic elements set in an environment of ecological sustainability. A breadth of vision is required to comprehend most problems. These factors are cultural, socio-economic and technical. The general and technical factors are complementary in a wider CCC framework.
- 4 *Circular causation*: Circular causation is very useful whether it be one of complete interdependence between variables or a circuitous-directional process of interaction. It is critical for the variables to interact through time in complex ways and for relationships to evolve through historical time. Circular causation also recognizes the need to transcend a narrow study of socio-economic institutions through linking specific institutional spheres (such as the financial system) to other spheres and relationships (such as households and governments). The broad socio-economic and the more technical economic factors interact in the CCC process.
- 5 *Cumulative causation*: Over long historical time, cumulative forces impact on the economy, as the linkages between major factors generate amplified and multiplied results from the initial changes. A change somewhere along the line is likely to have much greater effects than the initial ones. These effects are ongoing, usually not equilibrium-generating, and systemic in their impact. The interaction between the general and technical CCC factors stimulates more cumulative motion than when the socio-economic and technical are separate. Over time, transformations occur in the relationship between institutions and individuals.
- 6 *Reinforcing tendencies*: The secondary and tertiary changes will generally support the first, since various reinforcing effects operate in the economy. These reinforcing trends are of three main types. The first are the “trends to inequality” between regions and groups, as some move ahead while others are retarded. The second are “internal and external economies”, as externalities tend to be rampant, which reinforces inequality between areas and groups. The third are “spread effects”, as forward-looking impacts expand regions and groups, magnifying the initial changes.
- 7 *Path-dependence and lock-in*: One always needs to assess the relative importance of the reinforcing and counteracting forces. Irreversibility and path-dependence leads to the inability of the system to move back to the original equilibrium position (if there is one). The previous equilibrium is unlikely in the future, while a new equilibrium may not be forthcoming. The original changes – initial conditions – may set in motion forces conducive to the successful development and extension of a regime of accumulation (for



a nation or region). This regime may also lead to certain institutional and social relationships stimulating evolution throughout the social economy.

- 8 *Counteracting forces or contradictions*: The cumulative upward expansion or decumulation may be moderated by a whole series of forces, such as exogenous shocks, policy interference, negative externalities, industrial maturation and floors/ceilings to the cycle; plus changes in wages, population and enterprise profit. For instance, endogenous motion may eventually lead to the maturation of the regime of accumulation as the product cycle moderates. This may lead to lower growth as the habits and institutions are not conducive to structural changes in the regime. Other endogenous contradictions may also emerge, such as labour shortages as the numbers of rural workers decline; plus higher wages, material costs and interest rates.
- 9 *Waves of change*: Sometimes these changes are so great that they create major historical processes or waves of change. During these times parameters change, relationships between factors modify, and roles people play evolve. These waves of change generate potentialities that may result in new phases of development and growth over long historical time. Including both the general and technical factors in dynamic motion potentially increases the lags, and thereby stimulates periodicity and amplitude of the waves of change.
- 10 *Social foundations of development*: Since "the movement of the whole social system upward is ... what is mean[t] by development" (Myrdal 1968: 1868), the conditions necessary for development are broad and inclusive. The socio-economic institutions of capitalism may lead to an upward movement of the whole social system, while simultaneously generating inequality as other nations and groups are unable to benefit from the upward motion of CCC.

The social element of CCC starts by recognizing that CCC dynamics apply to ethnic, class and gender differences, as well as to national and regional dimensions of development. These cultural aspects of development and long-term transformation are critical to CCC. Whether we are looking at the problems of African Americans in an environment of a contradictory American creed of democracy, or problems of Asian development, this view of CC takes a long-term perspective of cultural transformation. Figure 6.1 outlines some of the major processes involved. This figure shows that poverty and underdevelopment – and its opposite, wealth and development – are historically associated with six main factors in complex ways. These factors include education and employment (human capital channels), trust and networks (social capital channels), prejudice and discrimination (asocial capital channels) habits, norms and mores (cultural capital channels), nutrition and psyche (health capital channels), as well as income and wealth (financial capital channels). These six forms of capital – human, social, asocial, cultural, health and financial – help to determine and in turn are influenced by the processes associated with poverty and underdevelopment (and its opposite, material progress and development) (O'Hara 2001). The critical thing here is interdependency: factors interact in multiple and complex

ways, impacting upon poverty and underdevelopment. One needs to scrutinize their realistic interaction through time via qualitative and quantitative explanations. As Myrdal recognized, linkages are circular (in complex ways), and their impact through long historical time tends to be cumulative (Myrdal 1944).

Class, gender, ethnicity, nations, regions and even organizations that are able to accumulate various forms of capital tend to develop and accumulate in an ongoing fashion. The groups that lag behind fall away relatively and perhaps absolutely. The circular and cumulative interactions of factors affect poverty and underdevelopment, as well as material progress and development. Cultural, institutional, political and even psychological factors play a role in economic processes as cybernetic and feedback interactions come into play. Both qualitative and quantitative elements are important in this full array of interdependencies.

On the one hand, areas or groups can surge ahead as they advance technologically, knowledge and skills accumulate, networks and organizations are formed, progressive habits and norms come into play, while nutritional and psychological factors improve (“spread effects”). Other areas lag behind because they lack resources, path-dependent expansions and institutional innovations; and these disadvantages encourage discrimination and prejudice, such as in the case of the African Americans versus white Anglo-Saxon US citizens; or Asian nations versus the impressions of citizens in advanced nations (in the 1960s). One group or area may have a cumulative upswing while the other experiences a vicious circle as one plane of living expands while the other is inhibited. The cumulative process will tend in this way to generate greater inequalities (Myrdal 1957: 12).

There could emerge, of course, counteracting forces to these specific cumulative ones, such as “exogenous factors”, policy changes and so on, which may have opposing impacts. As Myrdal (1957: 13) noted, however, these “backwash” factors are unlikely to propagate equilibrium tendencies, instead stimulating cumulative forces in a different direction. While to some degree, for instance, government policies “cannot change folk ways”, as Myrdal noted, they can

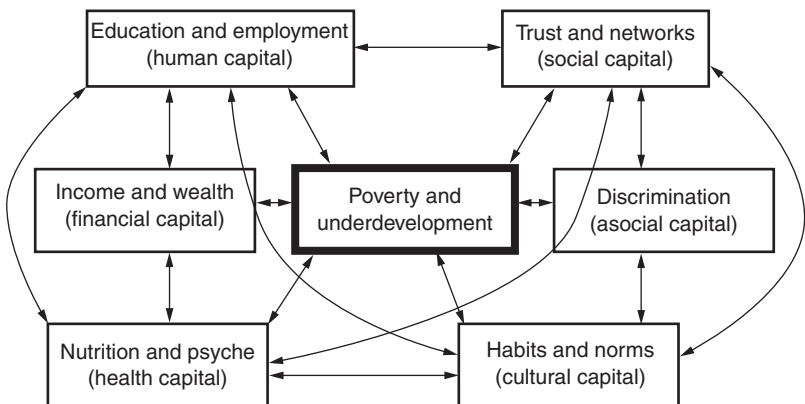


Figure 6.1 Cultural and socio-economic CCC dynamics.

condition them, moderate them, generating alternative folkways or changes in existing ones. Either way, the path-dependent process continues, perhaps in multiple directions and dimensions, towards more complex forms.

In Figure 6.2 we simplify the cultural and socio-economic aspects of CCC by the use of a four-segment quadrant. This quadrant illustrates a simple upward movement for a nation (and dominant group), while also illustrating that a minority group is unable to benefit to the same degree as the dominant group due to low (relative?) levels of resources, plus discrimination and prejudice. Segment A shows that human capital and income are positively related, as are social networks and income in Segment B. Segment C shows that discrimination declines when general networks rise, while Segment D illustrates that discrimination declines with higher levels of general human capital.

Linking the original curves generates equilibrium results (by a fluke). When change occurs, such as an upward movement in the income/networks nexus, from  $Y/N_0$  to  $Y/N_1$ , due to the relationship between variables circular and cumulative causation sets in. As networks stimulate more economies, income expands, which enables people to generate human capital, which reduces discrimination for those with capital and greater networks. Positive results occur as CCC motion generates a wave of upswing through the system. In this example, there are no endogenous contradictions except that those who were *not* included in the more effective networks and knowledge have fewer relative resources and perhaps more discrimination against them.

### *Technical foundations of long-term growth*

Demand–supply CCC is somewhat more technical, but still in continuity with the cultural and socioeconomic aspects of CCC. The technical details concern

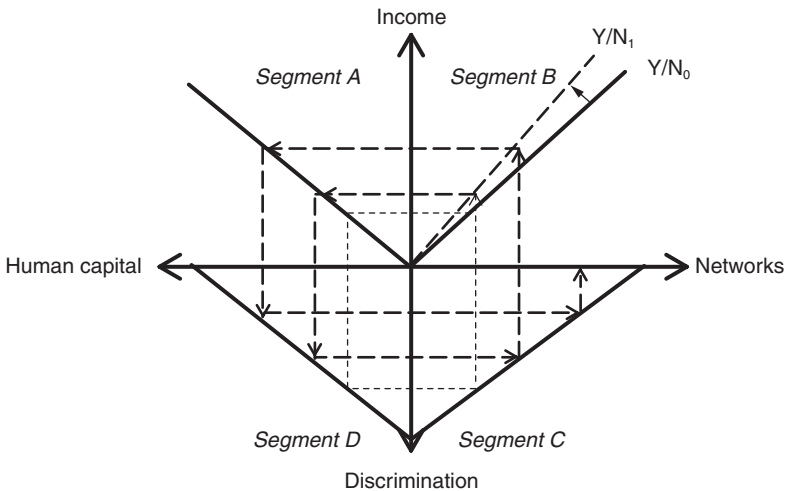


Figure 6.2 Cultural and socio-economic four-segment CCC quadrant.

the nature of the capitalist system, scaled economies, demand and productivity. In modern parlance, it is important to link demand–supply dynamics with technological change and economies of scope as the model becomes more realistic, as is illustrated in Figure 6.3. This shows how demand is the core of the problem, and that it is interdependent with supply. A strong level of domestic demand by itself and in relation to other factors can help provide an environment where confidence is relatively high (and uncertainty low), thereby stimulating investment at a relatively high rate. Investment can generate productivity increases through economies of scale and learning by doing (LBD) (Verdoorn's law); with spatial agglomeration, infrastructure and communications externalities playing critical roles. Demand and supply are thus interdependent, rather than the usual independent (supply) and dependent (demand) analysis of orthodoxy. A strong level of domestic income spurs productivity, which stimulates net exports, especially if the system of international finance is stabilizing and productive rather than overly speculative.

Critical here is the level of world income, which if high when uncertainty is relatively low can stimulate global trade. If the global environment has low levels of uncertainty, governments are cognizant of the need for productive public investment, and global finance stimulates relatively balanced systems of payment, then global exports can expand. This in turn stimulates domestic demand and investment, and through successive movements of the circuit also innovation and productivity, and so on *ad infinitum*. Nations with the leading sectors can especially undergo high levels of growth and prosperity. The circular and cumulative workings of capitalism can stimulate waves of upswing, as well as uneven development between the leading and underdeveloped nations. These circular and cumulative dynamics do not simply produce growth in the centre and uneven development in the world. The workings of *endogenous contradictions* can lead core nations to undergo maturation through lock-in of specific regimes of accumulation that are incapable of evolving into higher forms; or through declining underemployment as rural labour dries up.

Once the cultural and social foundations are developed the more technical economic aspects can expand; and *visa versa*. If the nation or region in question is able to expand from agriculture and mining through to highly productive

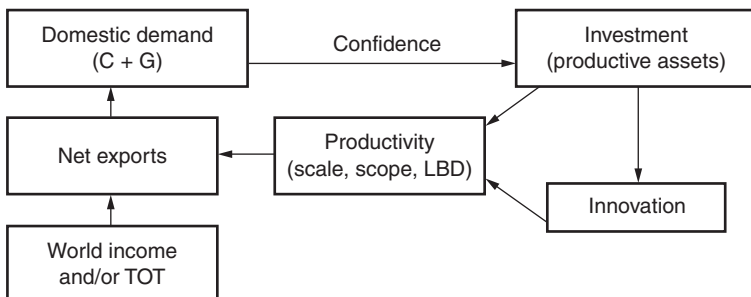


Figure 6.3 Kaldorian CCC dynamics.

manufacturing, transport, communications and utility sectors, economies can be generated through scale and scope. With relatively low levels of uncertainty, higher investment can stimulate the creation of a new regime of accumulation. This new regime creates winners and losers, but if world income is at high levels, the losers may be less numerous than the winners due to an expansion of productive world investment. Over time, though, the regime may dissipate as anomalies evolve.

Figure 6.4 isolates core elements of the process through a simple four-segment quadrant. With no change in the parameters, equilibrium may prevail along existing linear curves, such as  $P/Y_0$  (by a “fluke”). When a new accumulation regime emerges through a new investment/productivity dynamic, a new productivity/income curve arises in Segment B (from  $P/Y_0$  to nonlinear curve  $P/Y_1$ ). The consequent greater investment initially generates economies of scale with higher productivity and greater world income. Higher world income increases exports, which expands income, as multiple rounds of circular and cumulative dynamics ensue.

However, it is important to illustrate some contradictions in the model. The new regime of accumulation introduced with the nonlinear curve  $P/Y_1$  has an area that generates cumulation, but also an area of decumulation as the circle traverses in Segment B. There could be multiple causes of this contradictory motion. A lock-in of technology may occur if the regime of accumulation fails to adjust to a more viable one as maturation occurs (due to established industrial habits and norms). Another is a decline in levels of underemployment as rural

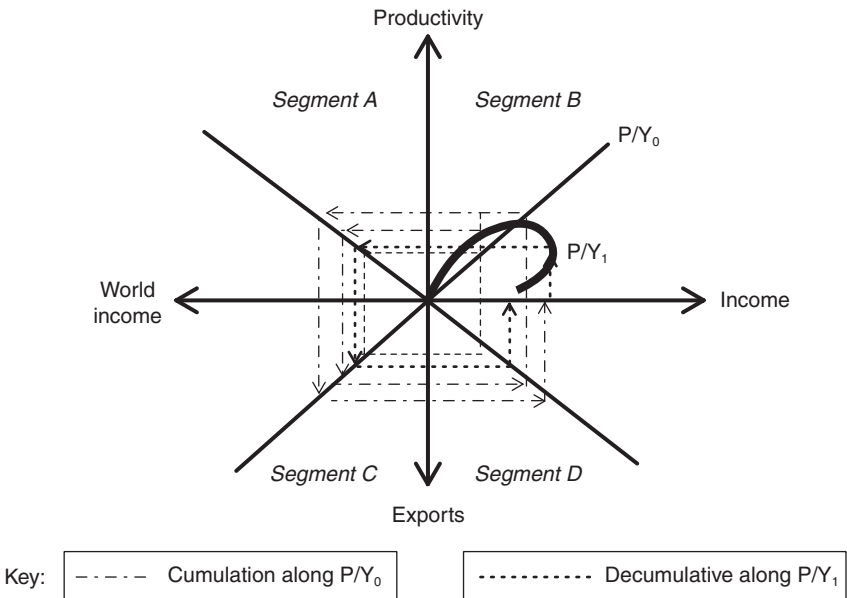


Figure 6.4 Demand-supply four-segment CCC quadrant with endogenous contradictions.

labour supply is exhausted through the movement to industry. These are the sorts of factors discussed by Kaldor (1966) and Eatwell (1982) in their accounts of Britain's economic malaise.

These simple figures and quadrants show how Myrdalian CCC can be formally linked with Kaldorian CCC to enhance growth and development theory. Myrdalian social development is both a pre-requisite and co-requisite for Kaldorian CCC; and visa versa. The explanatory power of CCC increases when the two are linked. The two CCC frameworks complement each other.

### Examples of CCC in the modern world: the schools of heterodox political economy

This final main section provides a further example of the workings of CCC in the modern world through an investigation of linkages between the schools of heterodox political economy (HPE) (see O'Hara 2008). Applications of this CCC model to the re-emergence and development of HPE can take the form of a full and a short model. The full model is illustrated in Figure 6.5. This CCC relationship is shown in shorthand form in Figure 6.6. Here the critical facet is the emergence and historical development and reproduction of contradictions (CR), as discussed above. The contradiction of the disembodied economy and destructive creation provide the historical theoretical and policy foundations of heterodoxy (HF). These foundations include the creative works of Marx, Veblen, Gilman, Schumpeter, Keynes and their followers, including Myrdal and Kaldor. This then enables general heterodox themes and ontology (GH) as well as individual schools concepts and concerns (IS) to develop. A key factor is the role of creative and proactive (CP) individuals (I), networks (N) and organizations (O), which primarily produce a productivity structure (PS) of publications (p), teaching (t), and socio-political (s) activities. This then links into the ongoing contradictory reproduction of heterodoxy as the circuit revolves in an ongoing fashion

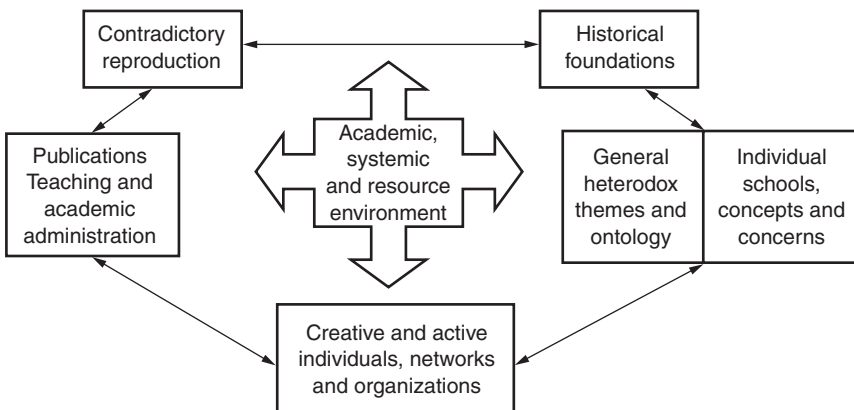


Figure 6.5 The circuit of HPE reproduction.

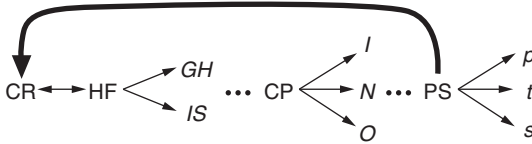


Figure 6.6 CCC relationship in shorthand form.

through numerous rounds of amplitude. Always impacting upon the circuit is the academic, systemic and resource environment.

Out of this set of influences, some discussion is necessary about the dual forces stimulating general heterodox themes and ontology as well as the concepts and concerns of the individual schools. The general heterodox themes are crucial for the emergence of concepts and principles promoting metamorphosis and evolution. Without these general concepts, little commonality could be shown to exist. For instance, heterodoxy tends to have a relatively uniform ontology that focuses on open systems, relationships between groups and individuals, and historical forces. Their analysis is realistic rather than *purely* technical and abstract. They tend to concentrate on circular processes linked with cooperation and competition, and they tend to be concerned with the rate of profit or surplus for capitalist enterprise. They centre on institutions and organizations set within the framework of individual action. All of them look at the forces of reproduction in one form or another, endogenous processes, instability and cycles, path-dependence and hysteresis; which themselves have circular and cumulative dynamics.

At the same time, there are concepts that link various schools, but not all. These are forces halfway between the specific concerns of individual schools and general concepts of heterodoxy. For instance, endogenous money and the financial instability hypothesis are core elements of Post Keynesian, institutional and Marxian thought. Segmented labour markets and class, gender and ethnicity are key areas of radical political economy (in between Marxism and institutionalism), feminism and institutionalism. Domestic labour in the past has been a research area of radical political economy and radical feminists. Cultural and social factors are becoming more general, but especially concern research by institutionalists, feminists and social economists.

Then there are the special concepts associated with the specific schools of thought. These are the areas of specialization enabling the promotion of deep conceptualization. For instance, Marxists have consistently been focusing on class, labour power, surplus value, circuit of social capital, exploitation and the mode of production. Institutionalists have been concentrating on conspicuous consumption and emulation; the instrumental and ceremonial functions of institutions; plus minimal dislocation and ceremonial encapsulation. Central areas of Post Keynesian thought include uncertainty, effective demand, liquidity preference, prospective yield and supply price, as well as the balance of payments constraint.



Schumpeterian and evolutionary scholars have been concentrating on the different forms of innovation, creative destruction, Schumpeterian competition, complexity and emergence, as well as on novelty and niches. Feminists have been developing a unique explanation for gender, caring labour, patriarchy, affirmative action and comparable worth, the double day, feminization of poverty and the glass ceiling. International and development scholars have been working through issues of uneven development, core and periphery, capabilities and the Prebisch-Singer hypothesis. Ecological political economists have been developing critical concepts and processes such as ecology and nature; entropy and negentropy; the precautionary approach; global warming and species extinction; the steady state economy; plus strong and weak sustainability.

These multiple forces of general heterodox concerns, interactions between some schools and specific schools of thought have been ongoing throughout history, including during the resurgence of political economy that has continued over the past 40 years. All three forces are necessary: general concepts, linkages between some schools, as well as specialization. The general concepts enable the schools to work together, developing broad principles of inquiry. Linkages between some schools of thought establish broader research programmes. The specialization enables the promotion of greater depth of more specific concepts. Through time transformation has been occurring, while innovations are instigated and new themes emerge, and many of the specific concepts of the respective schools have been shared. The circular and cumulative forces are complex, multifarious, ongoing, evolutionary and involving phases of metamorphosis. There is no finality, evolution rolls through over time, and blind drift operates, leading mostly to cumulative motion while affecting the general edifice of heterodoxy.

## **Conclusion**

This chapter has examined key aspects of circular and cumulative dynamics through works in the Myrdalian and Kaldorian traditions. There are many similarities, while the differences are complementary. Myrdal concentrated on the social foundations of development, while Kaldor scrutinized the technical foundations of long-term growth. Both are necessary for a fully fledged scrutiny of the development and growth process. Linking cultural-socio economic aspects of CC with technical demand-supply, CCC advances the explanatory power of political economy. Furthermore, CCC is capable of application to all major real world processes, including the recent development of heterodox political economy itself, involving the major schools and trends. CCC is associated with the circuit whereby heterodoxy is influenced by the contradictions of society, while stimulating general and specific themes and concepts, via creative individuals, networks and organizations, which in turn generate an array of publications, teaching and socio-political activities. The circular process continues through multiple interactive rounds of reproduction and evolutionary-cumulative change.

## Note

- 1 An earlier and much shorter version of this chapter was published in the *Journal of Economic Issues*, June 2008. I wish to thank Rick Adkisson, Sebastian Berger, Wolfram Elsner, Mathew Forstater and Vicky Taggart for their advice and assistance.

## References

- Argyrous, G. (2001) "Setterfield on Cumulative Causation and Interrelatedness: A Comment", *Cambridge Journal of Economics*, 25: 103–8.
- Berger, S. (2008) "Circular Cumulative Causation (CCC) à la Myrdal and Kapp – Political Institutionalism for Minimizing Social Costs", *Journal of Economic Issues*, 42, 2: 357–65.
- Eatwell, J. (1982) "The Principle of Cumulative Causation", ch. 3 of Eatwell *Whatever Happened to Britain? The Economics of Decline*, London: Duckworth.
- Fagence, M. (1980) "The Provision of Social and Other Services in Areas Experiencing Depopulation", *Social Policy and Administration*, 14, 2: 91–114.
- Gordon, D. M. (1991) "Kaldor's Macro System: Too Much Cumulation, Too Few Contradictions", in E. J. Nell and W. Semmler (eds) *Nicholas Kaldor and Mainstream Economics*, New York: St Martin's Press.
- Hawley, A. H. (1979) "Cumulative Change in Theory and History", in A. H. Hawley (ed.) *Societal Growth: Processes and Implications*, New York and London: Macmillan.
- Higgins, B. and Savoie, D. J. (1995) *Regional Development Theories and their Application*, New Brunswick: Transaction Publishers.
- Hodgson, G. M. (1989) "Institutional Rigidities and Economic Growth", in T. Lawson, J. G. Palma and J. Sender (eds) *Kaldor's Political Economy*, London: Academic Press.
- Humphrey, T. M. (1990) "Cumulative Process Models from Thornton to Wicksell", in D. E. Moggridge (ed.) *Perspectives on the History of Economic Thought, Volume IV: Keynes, Macroeconomics and Method*, Aldershot and Brookfield: Edward Elgar.
- Kaldor, N. (1966) *Causes of the Slow Rate of Economic Growth in the United Kingdom*, Cambridge, UK: Cambridge University Press; reprinted in N. Kaldor (ed.) (1978), *Further Essays on Economic Theory*, New York: Holmes & Meier.
- (1970) "The Case for Regional Policies", *Scottish Journal of Political Economy*, 27, 3: 234–63; reprinted in N. Kaldor (ed.) (1978), *Further Essays on Economic Theory*, New York: Holmes & Meier.
- (1972) "The Irrelevance of Equilibrium Economics", *Economic Journal*, 82, 328: 1237–55. reprinted in F. Targetti and A. P. Thirlwall (eds) (1989), *The Essential Kaldor*, New York: Holmes & Meier.
- (1975) "What is Wrong with Economic Theory", *Quarterly Journal of Economics*, 89, 3: 347–57.
- (1980) *Economics Without Equilibrium*, Armonk, NY: M. E. Sharpe.
- Lawson, T., Palma, J. G. and Sender, J. (eds) (1989) *Kaldor's Political Economy*, London: Academic Press.
- McCombie, J. S. L. and Thirlwall, A. P. (1994) *Economic Growth and the Balance of Payments Constraint*, New York: St Martin's Press.
- Myrdal, G. (1939; English edn 1965) *Monetary Equilibrium*, New York: Augustus M. Kelley.
- (1944) "A Methodological Note on the Principle of Cumulation", in G. Myrdal *An American Dilemma: The Negro Problem and Modern Democracy*, New York and

- London: Harper & Row; reprinted in G. Myrdal (1958), *G. Myrdal Value in Social Theory: A Selection of Essays on Methodology*, London: Routledge & Kegan Paul.
- (1957) “The Principle of Circular and Cumulative Causation”, ch. 2, in G. Myrdal *Economic Theory and Under-developed Regions*, London: Methuen & Co.
- (1968) “Appendix 2: The Mechanism of Underdevelopment and Development and a Sketch of an Elementary Theory of Planning for Development. 1. Circular Causation”, in G. Myrdal *Asian Drama: An Inquiry into the Poverty of Nations*, Vol. 3, Harmondsworth: Penguin Books.
- O’Hara, P. A. (2000) *Marx, Veblen and Contemporary Political Economy: Principles and Unstable Dynamics of Capitalism*, Cheltenham and Northampton: Edward Elgar.
- (2001) “Wealth Distribution and Global Inequality in the Multiple Capital Paradigm”, *International Journal of Human Development*, 1, 1: 121–40.
- (2007a) “Principles of Institutional-evolutionary Political Economy – Converging Themes from the Schools of Heterodoxy”, *Journal of Economic Issues*, 41, 1: 1–42.
- (2007b) “Heterodox Political Economy Specialization and Interconnection – Concepts of Contradiction, Heterogeneous Agents, Uneven Development”, *Intervention Journal of Economics*, 4, 1: 99–120.
- (2008) “Can the Principles of Heterodox Political Economy Explain Its Own Re-emergence and Development?”, *On the Horizon*, 16, 4: 260–77.
- Pini, P. (1995) “Economic Growth, Technological Change and Employment, Empirical Evidence for a Cumulative Growth Model with External Causation for Nine OECD Countries: 1960–1990”, *Structural Change and Economic Dynamics*, 6: 185–213.
- Setterfield, M. (1997) “History Versus Equilibrium and the Theory of Economic Growth”, *Cambridge Journal of Economics*, 21: 365–78.
- (2001) “Cumulative Causation, Interrelatedness and the Theory of Economic Growth: A Reply to Argyrous and Toner”, *Cambridge Journal of Economics*, 25: 107–12.
- Skott, P. (1999) “Kaldor’s Growth Laws and the Principle of Cumulative Causation”, in M. Setterfield (ed.) *Growth, Employment and Inflation: Essays in Honour of John Cornwall*, New York: St Martin’s Press.
- Tarkowski, J. (1988) “The Polish Crisis and Myrdal’s Model of Cumulative Causation”, *Political Studies*, 36, 3: 463–74.
- Toner, P. (1999) *Main Currents in Cumulative Causation – The Dynamics of Growth and Development*, New York: St Martin’s Press.
- (2001) “History versus Equilibrium and the Theory of Economic Growth, by Mark Setterfield: A Comment”, *Cambridge Journal of Economics*, 25: 97–102.

# 7 Circular cumulative causation à la Myrdal and Kapp<sup>1</sup>

*Sebastian Berger*

## Introduction

This chapter reconstructs the CCC from the writings of Myrdal and Kapp to explore the unique characteristics of this *key* concept of institutional economics. Moreover, it demonstrates the CCC's implications for *political* institutionalism. Incorporating new unpublished material from the Kapp Archive, namely the Myrdal–Kapp correspondence and Kapp's CCC lecture notes, I provide insights about the cooperation between the two economists and about Kapp's conceptual understanding of CCC. In addition, important differences to Veblen's CC<sup>V</sup> and Kaldor's CC<sup>K</sup> are pointed out to underline the CCC's uniqueness and its significance for institutional economics.

## Origin, meaning and significance

Myrdal formulated the CCC in Appendix 3 of *American Dilemma – The Negro Problem and Modern Democracy* (1944) for the first time, using it as a research hypothesis for the circular (self-reinforcing) causation between prejudices, institutions, and poverty. This triggers a vicious circle or “cumulative effect” of increasing inequalities, and poverty. Myrdal defined two distinct elements of the CCC (i.e. circular causation and its cumulative effect) in *Asian Drama – An Inquiry into the Poverty of Nations* (1968):

[c]ircular causation will give rise to a cumulative movement only when [...] a change in one of the conditions will ultimately be followed by a feed-back of secondary impulses [...] big enough not only to sustain the primary change, but to push it further. Mere mutual causation is not enough to create this process.

(Myrdal 1968: 1875)

Elsewhere he had formulated that “because of such circular causation a social process tends to become cumulative and often to gather speed at an accelerating rate” (Myrdal 1957: 13).

Kapp analyzed the circular cumulative causation of social costs since the 1940s and systematically elaborated the CCC's significance for the integration

of the social sciences in the 1950s (Kapp [1950] [1963] 1977c: 23, 25; 1961: 183, 187–8). Kapp considered the CCC to be *the* key concept of institutional economics (Kapp 1977b, 1968), and writes to Myrdal:

I read the Appendix [galley proofs of Appendix 2 of *Asian Drama*] with great profit; I am very much impressed [...] I think we now have a synthesis of an analysis in the sense of a theoretical framework and a system of tools for the study of the underdeveloped world [...] your tools go much beyond the underdeveloped countries and [...] a good deal of your new framework of analysis could be fruitfully applied in the study of developed countries.

(November 19, 1967, Unpublished Manuscript III)

CCC derives its status of a key concept from the fact that it contains *the* main antithesis to the mechanistic analogy and stable equilibrium of the social and economic system. As such it denies a necessary ameliorative trend in development, rejects stage theories, and anticipates the danger of poverty, and societal crisis. CCC may be considered to be the first scientific application of the ancient idea of the self-reinforcing vicious circle to socioeconomic problems. However, CCC is no doctrine of hopelessness because vicious circles can be broken, virtuous circles are possible, and a cumulative process also calls forth counteracting forces (Myrdal 1944: 1065; 1957: 35; 1968: 1857, 1859; Kapp, Unpublished Manuscript II).

As a precondition for self-reinforcing causation, CCC presupposes reciprocal causation and rejects the “*primum mobile*” causation theory (Kapp 1961: 188). Thereby, the CCC takes a stand in an old philosophical debate about cause and effect. For instance, Hegel was convinced that the reciprocal relation of cause and effect is the next truth that science will discover, whereas Schopenhauer stated that an effect cannot be the cause of its cause. In addition, Marx saw the development of the whole as being constituted of factors that interact with each other and with the whole. Regarding Marxian dialectics, Kapp argued that the CCC neither doubts the relevance of ideological nor material factors but that it rejects that one factor is *per se* exclusive and that the analysis can be restricted to it. Contradictions between material conditions and ideas are, however, possible (Kapp, Unpublished Manuscript II). Myrdal rejected a version of Marxian dialectics that attribute causal potency to the economic factor alone (Myrdal 1968: 1855–1905).

If social events and social change emerge in a process of reciprocal interaction between the elements of the system (i.e., within the inner structure), it is no longer adequate to attribute causal potency to an individual variable or impulse. Rather, the outcome (the event, the process) must be viewed as the result of the entire initial situation and the interaction process as well the basic properties of the total social structure.

(Kapp 1961: 188)

Therefore, attributing causal potency to the economic factor leads to only *seemingly* clear correlations. Denying the existence of a primary cause neither implies

a denial of the causal principle nor renders the search for relevant factors, their interdependence, or the direction of their change futile. The economic process is part of a larger social process and has to be analyzed as such. The CCC focuses on *all* relevant factors and rejects working with analytically closed models. The relevant factors can, of course, only be determined empirically in a given situation. Taking CCC's emphasis on interrelatedness seriously demands methodological interdisciplinarity and questions the autonomy of each social science. Thus the CCC, as applied in the context of economic planning, has focused on the conditions of the following categories of the social system that are by no means exclusive: (1) output and incomes; (2) conditions of production; (3) levels of living; (4) attitudes toward life and work; (5) policies; (6) institutions (Myrdal 1968). The CCC analysis confirmed the Veblenian account of the role of institutions (Kapp 1961, 1963b):

[c]ertainly the main resistance to change in the social system stems from attitudes and institutions. They are part of the inherited culture and are not easily or rapidly moved in either direction. Even in the very long run, attitudes and community institutions may stay much the same [...] the inertia of attitudes and institutions may be formidable [...] [and] the main reason why a "take off" may easily be abortive.

(Myrdal 1968: 1872)

Kapp even applied the CCC to circular causation between the open economic system and the environment, as well as circular causation in the environment (e.g. synergetic effects of different pollutants that disrupt the human environment in a cumulative process). In the light of the social system dynamics, exhibiting non-constant reciprocal interactions between the system and its parts that are often characterized by an uneven spread along the time axis, Kapp followed that the power of scientific analysis is limited and that it is doubtful whether a general theory or law can be expected of the social sciences. Consequently, Kapp as well as Myrdal asserted that there exists no math of social change, and both were quite skeptical of formal modeling (Kapp 1961: 188, 1965a; Unpublished Manuscript II; Myrdal 1968: 1866).

## Methodology

Methodologically CCC is a research hypothesis for capturing social dynamics and socio-ecological degradation. Kapp lists its main characteristics as follows: CCC (1) frames problems, (2) brings problems closer to solution, (3) requires an identification of relevant causal factors, (4) requires a causal analysis of real interaction relationships, (5) requires a systems view, (6) requires an analysis of temporal processes, (7) avoids teleology, the projection of ready-made meanings, relationships, results, and processes. Kapp even considered the CCC to have certain characteristics of a scientific "paradigm" with its own analytical apparatus, philosophical perspective, and hypothesis. However, to prevent

turning CCC into a dogma, Kapp even proposed to explore if there is indeed no single determining factor (Kapp 1961; Unpublished Manuscript II).

This methodology is compatible with substantive economics, as applied by Karl Polanyi (Polanyi 1957: 248; Berger 2008). By starting from a substantive notion of human needs and a corresponding social organization of livelihood, the substantive method aims at capturing, i.e. gaining a holistic understanding of economic phenomena in their specific local and historic context and their complex interrelations with the surrounding social and ecological systems. The cognitive function of substantive economics differs fundamentally from the formal approach in that it aims at an interaction with the actual world (Paul Tillich in Kapp 1961: 199), at staying close to empirical data, and at refraining from dogma. This method avoids the constant temptation to carry abstraction too far and to pursue formalization for its own sake. CCC is also compatible with the “real-typical” method or what has been called “Gestalt” theory in the tradition of Arthur Spiethoff (1953):

Their “real” character stems from what is empirically given. They are derived from the observed regularities of the social process, which are however, isolated from their historically unique and accidental context. What is retained for purpose of analysis are regularities as they are observed within the socio-cultural context.

(Kapp 1961: 198–9)

This way, CCC helps social science to stay on “an intermediate level of analysis between the idiographic approach of many historians and the level of abstraction at which pure theory or mathematics prefer to proceed” (Kapp 1965b: 8). By staying closer to the world of experience than pure theory and without adhering simply to an ideographic-descriptive approach, social theory follows the lead of its subject matter, in which causes and effects interact in a concatenation of reciprocal and cumulative interaction (Kapp 1965b: 8). Myrdal and Kapp also found that CCC shares important similarities with systems thinking that views the economic system as an open system in reciprocal interaction with the surrounding systems (Myrdal 1976: 215; Kapp 1976a). While noting important advantages of a “new” complex systems thinking, Kapp also cautioned that this could be but a first step to (1) view reality differently, (2) to order empirical findings in a new way, (3) to develop a language that fits the problems, and (4) to establish a basis for an adequate causal analysis which takes physical and institutional chains into account (Kapp 1972: 236; 1976a: 97). In a similar vein Kapp referred to John Dewey’s “pragmatist” methodology: “as Dewey has pointed out, [...] the formal concept of system acquires meaning and content only by making explicit the distinguishing characteristics of specific modes of association” (Kapp 1961: 103).

Also in more recent contributions, similarities between the CCC and general systems theory, system dynamics, and complexity theory have been noticed. For example, the CCC’s idea of self-reinforcing reciprocal causation is considered to



be an important building block of system dynamics' models working with "positive feedback loops" (Richardson 1991: 77) and the integration of institutional economics with formal models is proposed (Radzicki 2003: 133–73). However, this has called forth fundamental objections from institutionalists (Hayden 2006) that are similar to Kapp's skepticism regarding the application of formal-mathematical cybernetics in social analysis (Kapp 1961, Unpublished Manuscript I).

## Value premises

Applying the CCC to economic planning, Myrdal and Kapp both started from explicit value premises (Myrdal 1932, 1944: 1035–64; 1958; Kapp 1950, 1965a, 1973a). Desirable and possible positions of the system have to be justified normatively because they are neither automatic nor natural (Kapp, unpublished lecture notes). Both regarded it as a threat to science when research results are deemed "objective" while they are determined by *hidden* pre-analytical values.

The values made explicit by Myrdal are the democratic equality ideals of the Enlightenment that played a major role in the radical reforms of feudalism and the implementation of democratic institutions in Europe and the United States. This led to Myrdal's concern for poverty and increasing disparities between the rich and the poor that is similar to Kapp's concern for the fulfillment of basic human needs, i.e. his social minima approach (Kapp 1965a, 1965b). Aware of this link between both approaches, Myrdal expressed his sympathy in a letter to Kapp: "I have long felt that we are kindred souls and I am quoting you in a big book on South Asia [...] [I want to] thank you for your most interesting and inspiring article 'Social Economics and Social Welfare Minima'" (December 5, 1966, Unpublished Manuscript III).

Reflecting these value dispositions, the CCC serves as a research hypothesis to address vicious circles and the possibility of disruptive effects that are potentially threatening to human survival (Kapp 1976a; Berger and Elsner 2007). Kapp notes that the concept of "social reproduction" which was developed by the Physiocrats and was later adopted by Marx and Engels is of equal usefulness for the elaboration of hypotheses regarding defects and inefficiencies in the social system. CCC is also similar to development concepts developed by François Perroux ("domination effect": Perroux 1950, 1964: 32) and Johan Galtung ("center-periphery": Galtung 1975), although the CCC seems to be more generally applicable (Kapp [1973b] 1974: 132, 134–5).

The vicious circle of poverty is, according to Myrdal's CCC analysis, caused by inequalities in the (1) economic (e.g. income inequalities, including the unequal distribution of land), (2) social (e.g. lack of social mobility, including unequal educational opportunities), and (3) political (e.g. unequal participation in the political process, including qualitatively defective administration) realms within poor countries as well as their unequal trade relations with industrialized countries. In self-reinforcing circular causation these inequalities increase poverty dramatically in the already poor regions. In other words, these

inequalities are the cause and effect of poverty (Myrdal 1970; Kapp 1973a). Poverty as a form of insufficient satisfaction of basic human needs is a substantive indicator of social costs (Kapp 1963a).

Complementing the cause–effect analysis with a normative means–ends dimension, Kapp and Myrdal proposed “*political*” economics that is concerned with economic planning in real terms to guarantee social minima and minimize social costs (Myrdal 1960, 1968: 1879; Kapp 1965b, [1973b] (1974); Berger and Forstater 2007). Kapp argued that the open system character of the economy requires a new approach:

In short, as soon as the open character of economic systems is fully realised the formulation of social goals and objectives and the problem of collective choices can no longer be avoided. Such objectives and choices with respect to the maintenance of dynamic states of ecological and economic balance for the maintenance and improvement of the conditions of social and individual existence (quality of life) must become the point of departure of normative science of economics [...] the new task of economics would be to elucidate the manner in which collectively determined social goals and objectives could be attained in the most effective and socially least-costly manner.

(Kapp 1976a: 101–2)

Political institutionalism deals with policy-making aiming at the reduction of social costs (Kapp [1971] 1983). Kapp argued that the political process has to generate priorities in the light of defined human needs, i.e. “quality of life.” He believed in the scientific contribution to define the fundamental requirements of human life and survival as an integral part of a constellation of societal goals. The value of human needs could be objectified and transformed by scientific inquiry into social minima and maximum tolerance levels (e.g. for environmental pollution).

### **Veblen’s historical-philosophical CC<sup>V</sup>**

It has been argued that the CCC was derived from American institutionalism, i.e. Veblen’s concept of “cumulative causation” (Argyrous and Sethi 1996: 485) and that the two are more or less identical (e.g. Mayhew 2001: 243). Building on previous research (Berger and Elsner 2007) this chapter argues that Veblen’s concept of “cumulative causation” which he uses interchangeably with “cumulative change” CC<sup>V</sup> differs from Myrdal’s CCC in several ways. Regarding the origin, it has been argued that Veblen derived the main idea for the CC<sup>V</sup> from Herbert Spencer’s “historical-evolutionary” teachings (Jennings and Waller 1998: 196–8) or from Peircian philosophy (Hall and Whybrow: see Chapter 11, this volume). Overall it seems that Veblen’s CC<sup>V</sup> is more historical and philosophical. Applying the CC<sup>V</sup> mainly to “the sequence of change in the methods of dealing with the material means of life” (Veblen 1898: 387), Veblen described

the economic life history as a “cumulative process of adaptation of means to ends that cumulatively change as the process goes on” (Veblen 1898: 391). According to Veblen, the main characteristic of an evolutionary economist is that “he insists on an answer in terms of cause and effect [...] the notion of cumulative causation” (Veblen 1898: 377). It seems clear that Veblen’s concept captures how past conditions influence the present, how the present evolved out of the past, and how the present currently continues to evolve. Veblen’s CC<sup>V</sup> also implies that societal evolution has no specific end and that it is self-propagating, i.e. endless and ever continuing. The CC<sup>V</sup> states that there is “no trend” in cumulative causation: “[The] scheme of blindly cumulative causation, in which there is no trend, no final term, no consummation” (Veblen 1907: 304). Veblen considered the CC<sup>V</sup> as a conceptual alternative to teleological or stage conceptions of socio-economic dynamics. Asserting that socio-economic evolution is not unidirectional; Veblen’s CC<sup>V</sup> underlines that it can take the form of progress as well as regress to earlier barbarian stages.

This rejection of imputing trends to social evolution does not mean that Veblen did not make the phenomenon of institutional inertia one of his main research areas. The latter is an early theory of path-dependence and “lock-in” that keep the socio-economic system on a certain evolutionary path for a considerable time period. This, of course, is similar to the idea of a tendency and Veblen related this to the human behavioral propensity for emulation that has to some degree a self-reinforcing effect. It seems that Veblen’s CC<sup>V</sup> rejects a pre-analytical (i.e. general and non-empirical) assertion of a trend, while allowing for the identification of tendencies and counter-tendencies in specific contexts and for limited periods of time.

This shows how Veblen’s definition and application of CC<sup>V</sup> differ from Myrdal’s CCC. The former, per se, neither contains the idea of self-reinforcing causation nor the notion of a definite upward or downward movement of the social fabric. Hence the CCC cannot simply be read as a theoretical module already introduced by Veblen or identical to CC<sup>V</sup>. This explains why Myrdal did not refer to Veblen’s CC<sup>V</sup> in his “Remarks upon Receipt of the Veblen-Commons Award” (Myrdal 1976: 215). Against this background it is also warranted to rely on Myrdal himself, who remarked that he derived the concept from his earlier models in *Monetary Equilibrium* (1939) and names Knut Wicksell as a forerunner of the CCC (Myrdal 1944: 1065 fn. B; see also Wahid 2002: 85; Lundberg 1994: 426–30). After all, approaches dealing with “cumulative” effects and circular causation were common in England, Germany, and Sweden in the 1920s. Suffice to mention the work of Knut Wicksells on inflation and Myrdal’s work on the dynamics of savings and investment rates (e.g. Sandelin 1991: 186–9). In addition, the “Kiel School” worked on the causes of the trade cycle (Forstater 2003: 309; Krohn 1993) and Adolph Löwe had already conceptualized the trade cycle as “reciprocal causation in a circular flow” (Löwe 1935: 124–7):

Through the medium of the exchange process the primary cause or causes of the trade cycle, themselves arising from the social environment of the

exchange process, react upon that very environment. In the face of such reciprocal action between social and economic factors, it is hardly possible to interpret the true relation between economics and sociology as a one-sided dependence.

(Löwe 1935: 103)

Hence, instead of constructing a direct Veblenian heritage for all versions of CCC it seems more adequate to accept the more or less simultaneous emergence of related but not identical concepts. Recent studies have stressed the difference between the two understandings as well and have argued that Myrdal's CCC grew out of dealing with complex problems with a clear policy orientation (Angresano 1997: 85).

That said, the CCC is, nevertheless, compatible with Veblen's CC<sup>V</sup> research program for an evolutionary economics. Similarities between the two exist and have been noticed by Kapp: (1) the rejection of a tendency toward equilibrium, (2) viewing cultural and economic processes as unending, non-teleological, and moved by institutional inertia (Kapp [1950] [1963] 1977c: 25).

### Kaldor's "growth theoretical" CC<sup>K</sup>

In the literature on Kaldor's CC<sup>K</sup> some authors point out similarities with Myrdal's CCC regarding the role of the state, asymmetric trade relations, and evolution (Argyrous 1996: 110; Skott 1994: 119; Skott and Auerbach 1995; Toner 2001: 100–3). Similarities such as the rejection of equilibrium economics, the focus on unequal trade relations, and economic inequality are certainly important. However, it is a well-known fact that Kaldor's CC<sup>K</sup> also differs from the CCC in the Myrdal-Kapp tradition in several respects and has been noted, for example, by Setterfield who detects a unique Kaldorian tradition (Setterfield 2001: 109) and Skott who called Kaldor's approach "more traditionally economic" (Skott 1994). It seems worthwhile to elaborate on these differences a little further.

Kaldor's interpretation of Myrdal's CCC is mainly based on *Economic Theory And Under-developed Regions* (Myrdal 1957). In a sub-chapter called "The principle of cumulative causation," Kaldor wrote, "what Myrdal called the principle of 'circular and cumulative causation' [...] is nothing else but the existence of increasing returns to scale" (Kaldor [1970] 1978: 143). In another sub-chapter called "The theorem of endogenous and cumulative change," Kaldor refers to Allyn Young's theory that "with increasing returns change becomes progressive and propagates itself in a cumulative way" and asserts that "Myrdal [...] called this the 'principle of circular and cumulative causation'" (Kaldor [1972] 1978: 186).

Kaldor's interest in the economic mechanism of "increasing returns" led to this narrow interpretation of Myrdal's CCC. It is true that Kaldor generally conceded the importance of non-economic factors in the development process, such as the effects of governmental intervention and training on growth (Argyrous 2001: 105; Toner 2001: 99–101). However, in Kaldor's writings on CC<sup>K</sup> an

attempt to explain how these relevant factors of the social system are interrelated is missing. Although Kaldor's "interrelatedness" portrayed the economy as a "complex web of interrelations" (Argyrous 2001: 105) it focused mainly on the interdependencies of "economic" components, such as machinery and organizational structure (Toner 2001: 100). The fact that Kaldor's analysis of institutions was of secondary importance has also been noticed by Toner. Therefore, Toner's conclusion that Kaldor advanced CC<sup>K</sup> theory only applies to more narrow economic factors (Toner 1999: 115). Kaldor put back on the research map topics such as the balance of payment constraint to growth, the key role of manufacturing in development, and a broad conception of increasing returns. This more limited "economic" research interest may explain why he preferred the term "cumulative causation" over "circular causation." This does not necessarily imply that Kaldor was indifferent or hostile to Myrdal's broader approach, but it is precisely the latter's broader approach that makes for the CCC's unique potential for *institutional* economics.

Focusing *only* on "fundamental economic" principles to understand problems of unequal economic development while treating everything else as more or less "exogenous," and to derive policy conclusions from such an analysis is different from Myrdal's and Kapp's research interests. In the last instance the different research agendas also seem to be related to Myrdal's and Kapp's value orientation. Although Kaldor is clearly concerned about unequal economic growth, he does not explicitly address the importance of value premises, and does not propose political economics. Since Kaldor does not enter into a discussion of means and ends it can only be inferred that Kaldor considers industrial growth as a given goal. Seemingly, the proposed solution to the problem of unequal growth would be more industrial growth for those lagging behind. It seems to be implied that industrial growth (i.e. growth in manufacturing (factory system)) is paralleled with growth in general.

Myrdal and Kapp were critical of industrialization void of socio-ecological goals and criticized formal treatments of the development process as "empty" because they avoid "political" aspects. It seems that the notion of circular causation between many different variables in the social system opens the door to an analysis of the full scale of interrelated causes and effects of development problems. Such an analysis leads to a meaningful discourse of substantive issues, such as poverty, environmental disruption, vested interests, corrupt power elites, unequal educational opportunities, and unequal land distribution. For example, industrial growth has little meaning when a large part of the population remains poor and when environmental disruption contributes to the vicious circle of poverty. Hence, industrial growth without further detailed parameters is not necessarily the cure-all for the complex problem of unequal development. In fact, it is conceivable that there are problem contexts in which a solution may require less or no growth in manufacturing, or only a specific kind of manufacturing adequate to the problem. This issue is related to the method of measuring growth (i.e. the success of development efforts), and to CCC's value orientation. For example, indicators of exchange values (market prices) are arbitrary and do

not adequately reflect the status quo of the quality of life or success in the “community’s industrial arts.” On the contrary, they can conceal great disparities, social costs, and a dehumanization of social reality. Consequently, different kinds of growth, as well as the meaning and measurement of growth are part of the discussion in CCC’s value orientation. On the whole, the approaches of Myrdal and Kapp seem to be relatively closer to Veblen’s critical institutional analysis than to traditional growth economics.

## Conclusion

This chapter has shown how Kapp and Myrdal developed the CCC into *the* key concept of institutional economics. In addition, I pointed out the differences to Veblen’s CC<sup>V</sup> and Kaldor’s CC<sup>K</sup> evidencing that the CCC holds the fuller theoretical and conceptual potential for *institutional* economics.

## Note

- 1 An earlier version of this chapter was presented at the annual meeting of the Association for Evolutionary Economics in New Orleans, LA, January 4–6, 2008 and published in the *Journal of Economic Issues*, June 2008. I am grateful to the discussants Mathew Forstater and Edward Nell for their helpful comments.

## References

- Angresano, J. (1997) *The Political Economy Of Gunnar Myrdal: An Institutional Basis for the Transformation Problem*, Cheltenham and Northampton, MA: Edward Elgar.
- Argyrous, G. (1996) “Cumulative Causation and Industrial Evolution: Kaldor’s Four Stages of Industrialization as an Evolutionary Model,” *Journal of Economic Issues*, 30, 1: 97–119.
- (2001) “Setterfield on Cumulative Causation and Interrelatedness: A Comment,” *Cambridge Journal of Economics*, 25, 1: 103–6.
- Argyrous, G. and Sethi, R. (1996) “The Theory of Evolution and the Evolution of Theory: Veblen’s Methodology in Contemporary Perspective,” *Cambridge Journal of Economics*, 20, 4: 475–95.
- Berger, S. (2008) “Karl Polanyi’s and K. William Kapp’s Substantive Economics: Important Insights from the Polanyi–Kapp Correspondence,” *Review of Social Economy*, 66, 3: 381–96.
- Berger, S. and Elsner, W. (2007) “European Contributions to Evolutionary Institutional Economics: The Cases of Cumulative Circular Causation (CCC) and Open System Approach (OSA) – Some Methodological and Policy Implications,” *Journal of Economic Issues*, 41, 2: 529–37.
- Berger, S. and Forstater, M. (2007) “Towards Political Institutional Economics. Kapp’s Social Costs, Lowe’s Instrumental Analysis, and the European Institutional Approach to Environmental Policy,” *Journal of Economic Issues*, 41, 2: 539–46.
- Forstater, M. (2003) “Cumulative Causation à la Lowe: Radical Endogeneity, Methodology, and Human Intervention,” in G. Argyrous, M. Forstater and G. Mongiovi (eds) *Growth, Distribution, and Effective Demand – Alternatives to Economic Orthodoxy, Essays in Honor of Edward J. Nell*, London: M. E. Sharpe.

- Galtung, J. (1975) "Development From Above and the Blue Revolution: The Indo-Norwegian Project in Kerala," in *International Peace Research Institute*, Oslo: PRIO Publication No. 2–12.
- Hayden, G. F. (2006) "The Inadequacy of Forrester's System Dynamics Computer Programs for Institutional Principles of Hierarchy, Feedback, and Openness," *Journal of Economic Issues*, 40, 2: 527–35.
- Jennings, A. and Waller, W. (1998) "The Place of Biological Science in Veblen's Economics," *History of Political Economy*, 30, 2: 189–217.
- Kaldor, N. [1970] (1978) "The Case of Regional Policies," in N. Kaldor (ed.) *Further Essays on Economic Theory*, New York: Holmes & Meier.
- [1972] (1978) "The Irrelevance of Equilibrium Economics," in N. Kaldor (ed.) *Further Essays on Economic Theory*, New York: Holmes and Meier Publishers.
- Kapp, K. W. (1950) *The Social Costs of Private Enterprise*, Cambridge, MA: Harvard University Press.
- (1961) *Toward a Science of Man in Society – A Positive Approach to the Integration of Social Knowledge*, The Hague: Martinus Nijhoff.
- (1963a) "Social Costs and Social Benefits – A Contribution to Normative Economics," in E. v. Beckerath and H. Giersch (eds) *Probleme der normativen Ökonomik und der wirtschaftspolitischen Beratung*, Berlin: Duncker & Humblot.
- (1963b) *Hindu Culture, Economic Development and Economic Planning in India*, Bombay and London: Asia Publishing House.
- (1965a) "Economic Development in a New Perspective: Existential Minima and Substantive Rationality," *Kyklos*, 17, 1: 49–79.
- (1965b) "Social Economics and Social Welfare Minima," in T. K. N. Unnithan *et al.* (eds) *Towards a Sociology of Culture in India, Essays in Honor of Dr. D. P. Mukerji*, New Delhi: Prentice Hall of India.
- (1968) "In Defence of Institutional Economics," *Swedish Journal of Economics*, 70, 1: 1–18.
- (1972) "Umweltkrise und Nationalökonomie," *Schweizerische Zeitschrift für Volkswirtschaft und Statistik*, 108, 3: 231–49.
- (1973a) "Entwicklungspolitik in neuer Perspektive: Bemerkungen zu Gunnar Myrdals 'Politisches Manifest über die Armut in der Welt'," *Blätter für deutsche und internationale Politik*, 18, 3: 1–12.
- [1973b] (1974) "Environmental Indicators as Indicators of Social Use Value," in K. W. Kapp (ed.) *Environmental Policies and Development Planning in Contemporary China and other Essays*, Paris, The Hague: Mouton.
- (1976a) "The Open-system Character of the Economy and its Implications," in K. Dopfer (ed.) *Economics in the Future*, London: Macmillan.
- [1976b] (1977a) "Development and Environment: Towards a New Approach to Socio-economic and Environmental Development," in R. Steppacher, B. Zogg-Walz and H. Hatzfeld (eds) *Economics in Institutional Perspective – Memorial Essays in Honor of K. W. Kapp*, Lexington, MA, and Toronto: Lexington Books.
- (1977b) "The Nature and Significance of Institutional Economics," *Kyklos*, 29, 2: 209–32.
- (1950; 2nd edn 1963; 3rd edn 1977c) *The Social Costs of Business Enterprise*, Nottingham: Spokesman University Paperback.
- [1967] (1985) "Economics and Rational Humanism," in J. E. Ullmann and R. Preiswerk (eds) *The Humanization of the Social Sciences*, Lanham, MD, and London: University Press of America.



- (1971) "The Implementation of Environmental Policies, in: Development and Environment. Report and working papers of a panel of experts convened by the Secretary General of the United Nations Conference on the Human Environment, Founex, Switzerland, June 4–12, 1971," in J. E. Ullmann (ed.) (1983), *Social Costs, Economic Development and Environmental Disruption*, Lanham, MD: University Press of America.
- Unpublished Manuscript I "General System Theory and the Integration of Social Science," *Kapp Archive*, Basle (CH), Estimated date of manuscript, late 1950s.
- Unpublished Manuscript II "CCC lecture notes," *Kapp Archive*, Basle (CH).
- Unpublished Manuscript III "Kapp–Myrdal Correspondence," *Kapp Archive*, Basle (CH).
- Krohn, C. D. (1993) *Intellectuals in Exile – Refugee Scholars and the New School for Social Research*, Massachusetts: The University of Massachusetts Press.
- Löwe, A. (1935) *Economics and Sociology – A Plea for Co-operation in the Social-sciences*, London: Allen & Unwin.
- Lundberg, E. (1994) "Gunnar Myrdal's Contribution to Economic Theory: A Short Survey," in R. Henriksson (ed.) *Erik Lundberg Studies in Economic Instability and Change: Selected Writings Through Five Decades together with an Obituary by William J. Baumol*, Stockholm: SNS Förlag.
- Mayhew, A. (2001) "Human Agency, Cumulative Causation, and the State," *Journal of Economic Issues*, 35, 2: 239–50.
- Myrdal, G. [1932] (1976) "Das politische Element in der national-ökonomischen Doktrinbildung," Bonn-Bad Godesberg: Verlag Neue Gesellschaft.
- (1939) *Monetary Equilibrium*, London: W. Hodge & Company.
- (1944) *American Dilemma – The Negro Problem and Modern Democracy*, New York and London: Harper & Row.
- (1957) *Economic Theory and Under-developed Regions*, London: Gerald Duckworth.
- (1958) *Value in Social Theory – A Selection of Essays on Methodology by Gunnar Myrdal*, London: Routledge & K. Paul.
- (1960) *Beyond the Welfare State – Economic Planning in the Welfare States and its International Implications*, New Haven, CT: Yale University Press.
- (1968) *Asian Drama – An Inquiry Into the Poverty of Nations*, New York: Pantheon.
- (1970) *The Challenge of World Poverty – A World Anti-poverty Program in Outline*, New York: Pantheon Books.
- (1976) "Remarks upon Receipt of the Veblen-Commons Award," *Journal of Economic Issues*, 10, 2: 215–16.
- Perroux, F. (1950) "The Domination Effect and Modern Economic Theory," *Social Research*, 17, 1: 188–206.
- (1961; 2nd edn 1964) *L'économie du XXème Siècle*, Paris: Presses Universitaires de France.
- Polanyi, K. (1957) "The Economy as an Instituted Process," in K. Polanyi, C. M. Arensberg and H. W. Pearson (eds) *Trade and Market in the Early Empires – Economies in History and Theory*, Glencoe, IL: Free Press.
- Radzicki, M. J. (2003) "Mr. Hamilton, Mr. Forrester, and a Foundation for Evolutionary Economics," *Journal of Economic Issues*, 37, 1: 133–73.
- Richardson, G. P. (1991) *Feedback Thought in Social Science and Systems Theory*, Philadelphia, PA: University of Philadelphia Press.
- Sandelin, B. (1991) *The History of Swedish Economic Thought*, London and New York: Routledge.

- Setterfield, M. (2001) "Cumulative Causation, Interrelatedness and the Theory of Economic Growth: A Reply to Argyrus and Toner," *Cambridge Journal of Economics*, 25, 1: 107–12.
- Skott, P. (1994) "Cumulative Causation," in G. M. Hodgson, W. J. Samuel and M. R. Tool (eds) *The Elgar Companion to Institutional and Evolutionary Economics (Vol. 1)*, Aldershot: Edward Elgar.
- Skott, P. and Auerbach, P. (1995) "Cumulative Causation and the 'New' Theories of Economic Growth," *Journal of Post-Keynesian Economics*, 17, 3: 381–403.
- Spiethoff, A. (1953) "Pure Theory and Economic Gestalt Theory; Ideal Types and Real Types," in F. C. Lane and J. C. Riemersma (eds) *Enterprise and Secular Change – Readings in Economic History*, Homewood, IL: Richard D. Irwin, Inc.
- Toner, P. (1999) *Main Currents in Cumulative Causation – The Dynamics of Growth and Development*, New York: St Martin's Press.
- (2001) "History versus Equilibrium and the Theory of Economic Growth, by Mark Setterfield: A Comment," *Cambridge Journal of Economics*, 25, 1: 97–102.
- Veblen, T. B. (1898) "Why is Economics not an Evolutionary Science?," *The Quarterly Journal of Economics*, 12, 4: 373–97.
- (1907) "The Socialist Economics of Karl Marx and His Followers," *The Quarterly Journal of Economics*, 21, 2: 299–322.
- Wahid, A. N. M. (2002) *Frontiers of Economics – Nobel Laureats of the Twentieth Century*, Westport, CT: Greenwood Press.

# 8 Utilizing the social fabric matrix to articulate circular and cumulative causation for conceptual conclusions

*F. Gregory Hayden*

## Introduction

Scientists from many different fields of study independently derived common principles about systems that are usually referred to as general systems analysis (GSA), which can be applied for analysis with the social fabric matrix (SFM) (see Hayden 2006: 51–60, 94–106). The institutionalist Gunnar Myrdal used the term “circular and cumulative causation” (CCC) to describe principles he derived for the analysis of socioeconomic systems (1944: 1065–70; 1974: 719–32; 1978: 774–5). His findings regarding systems were consistent with GSA discovered in other fields such as physics, biology, and anthropology. The purpose here is to present a SFM in order to use the empirical content of a real-world socioeconomic system to derive conceptual conclusions about CCC/GSA concerns, and to comment on a current controversy regarding rule emergence. The SFM, digraph, and cellular description are completed for part of the Nebraska State system used to distribute state funds among local K-12 public schools. The Nebraska study combines the problem orientation of instrumentalism and the systems analysis of CCC/GSA. (The SFM elements from the complete study are available at the interactive SFM website <http://cba.unl.edu/academics/economics/sfm/>.)

The SFM approach to scientific analysis and policy evaluation allows for the knowledge base about values, social beliefs, institutions, attitudes, technology, and the ecological system to be assembled in such a manner to articulate the transactional relationships among the real-world components of those concepts in order to discover the system network and processes that define and guide the components. “The focus of the SFM is to provide a means to assist in the integration of diverse fields of scientific knowledge, utilize diverse kinds of information in order to describe a system, . . . evaluate policies and programs, and create social indicators for future monitoring” (Hayden 2006: 73). Myrdal emphasized that the coefficients of interrelations among the various parts and conditions of systems “usually are *unknown*, or our knowledge of this is *utterly imprecise*,” because of a “tremendous area of ignorance” and the adoption of concepts “that are not adequate to local reality” (1974: 730–1). The SFM provides an approach to reverse all three of those concerns.

To complete the SFM, the same components are arrayed in the same order as rows and columns of an adjacency matrix with the cellular information being the deliveries from row components to column components. Consistent with reality, the SFM does not impose the requirement that deliveries have a common denominator, and, consistent with open systems, it does not express system equilibrium. The edges between components in the SFM digraph (as displayed in Figure 8.1) represent the cellular deliveries between row and column components and deliveries between the environment outside the system and the components.

A claim is often made that CCC/GSA is flawed because it has no sound strategy for selecting and dividing variables when a whole system approach is to be utilized (see Bailey 1994: 65). The evidence for such a claim is to some extent because some system scientists have emphasized the conversion of system concepts into rarified abstractions rather than taking the concepts into the field to apply with the detail of reality. In addition, statisticians have convinced so many that their analytical strategy for the division of variables is legitimate when it ignores the system context. "A major problem with this analytical strategy is that focus is displaced from the systems viewpoint. Rather than focus on understanding human social groups in all their ramifications, one is focusing on narrow technical concerns" (Bailey 1994: 64). The CCC/GSA approach is concerned with explanation of the system. The strong commitment of statisticians to data-mining and technical manipulations does little to aid system explanation, and often detracts from it.

It would be distinctly preferable from the standpoint of systems theory to first describe the whole system ... instead of only a sample of persons abstracted from some system. Statisticians often speak of a population or universe from which their sample is drawn. The problem from a systems perspective is that this universe *may or may not coincide with some concrete system*. After the whole system is described, then the important variables affecting its operation should be described, with an emphasis on their relationships and interrelationships, including multiple interactions and nonlinear relationships.

(Bailey 1994: 64–5)

John Dewey's emphasis on a problem orientation used in combination with the SFM solves the concern about the inability to select variables and describe a concrete system. According to Dewey's problem orientation, the problem selected for research is to guide the selection of the system context, processes, and variables. The system can be made concrete with the SFM because it provides a comprehensive strategy for variable selection and definition in a concrete setting. The SFM is designed for researchers to complete the matrix with component variables and deliveries among the components found from the investigation of a real-world context.

## School aid system in Nebraska

The SFM for the system considered here is reported in digraph format in Figure 8.1. It is a system taken from a larger whole consistent with the problem selected. The main system for state school aid in Nebraska is titled the Tax Equity and Educational Opportunities Support Act (TEEOSA). The emphasis of the SFM study was to express the rules of TEEOSA that determine how much money is distributed by the state to each of the 257 school districts, and to explain the relationships among the different rules. The SFM cell descriptions are algebraic articulations of TEEOSA rules. These rules result in a continuous algebraic formula for each district that is 600 pages long (abridged to 138 pages in the website). The rules were discovered and converted to algebra by reviewing the laws, regulations, and standard operating procedures for each term and by interviewing and observing personnel in the departments and divisions responsible for calculating and programming the numbers that represent the amount of money a district is to gain or lose from each term in the TEEOSA process. Government departments are the producers of the numbers which are the work product in this case. Formula terms require the institutional organizations to coordinate work and data from the local, state, and federal government. The numbers produced and the use of those numbers – that is, the integrated work completed by these organizations – are determined by rules from the Nebraska Senate (Unicameral) and Nebraska Courts, as indicated in Figure 8.1.

As depicted in Figure 8.1, legislative and court decisions are delivered to the Nebraska Department of Education (NDE), which directs its divisions to follow the rules and to coordinate with other divisions and other departments in order to fulfill the regulations and requirements of the rules. The directed edges from NDE to the other components represent rules delivered to those components on what and how to calculate particular aspects of the overall formula, what data to use, from where to acquire it, and to what groups the calculations should be reported. The number preceding each component in Figure 8.1 is its SFM row and column number. Institutional components 32 through 39 are given the name that corresponds to TEEOSA rules that designate the components' activities. They are as follows:

- *Net Option Students*: Parents have the option of sending children to a school district in which they do not reside. Net Option Students for a district are the difference between the number of option students received from another district (opting in) and the number of a district's option students going to another district (opting out). These students have already been counted once for the districts, but are counted again as Net Option Students. According to the rules, the numbers of Net Option Students compiled by the interaction of local districts and state personnel are directed to those who calculate net option funding (Neb. Rev. Stat. § 79–233).<sup>1</sup>
- *Net Option Funding*: Net Option Funding is a long complicated formula term that is repeated numerous times throughout the larger formula. As

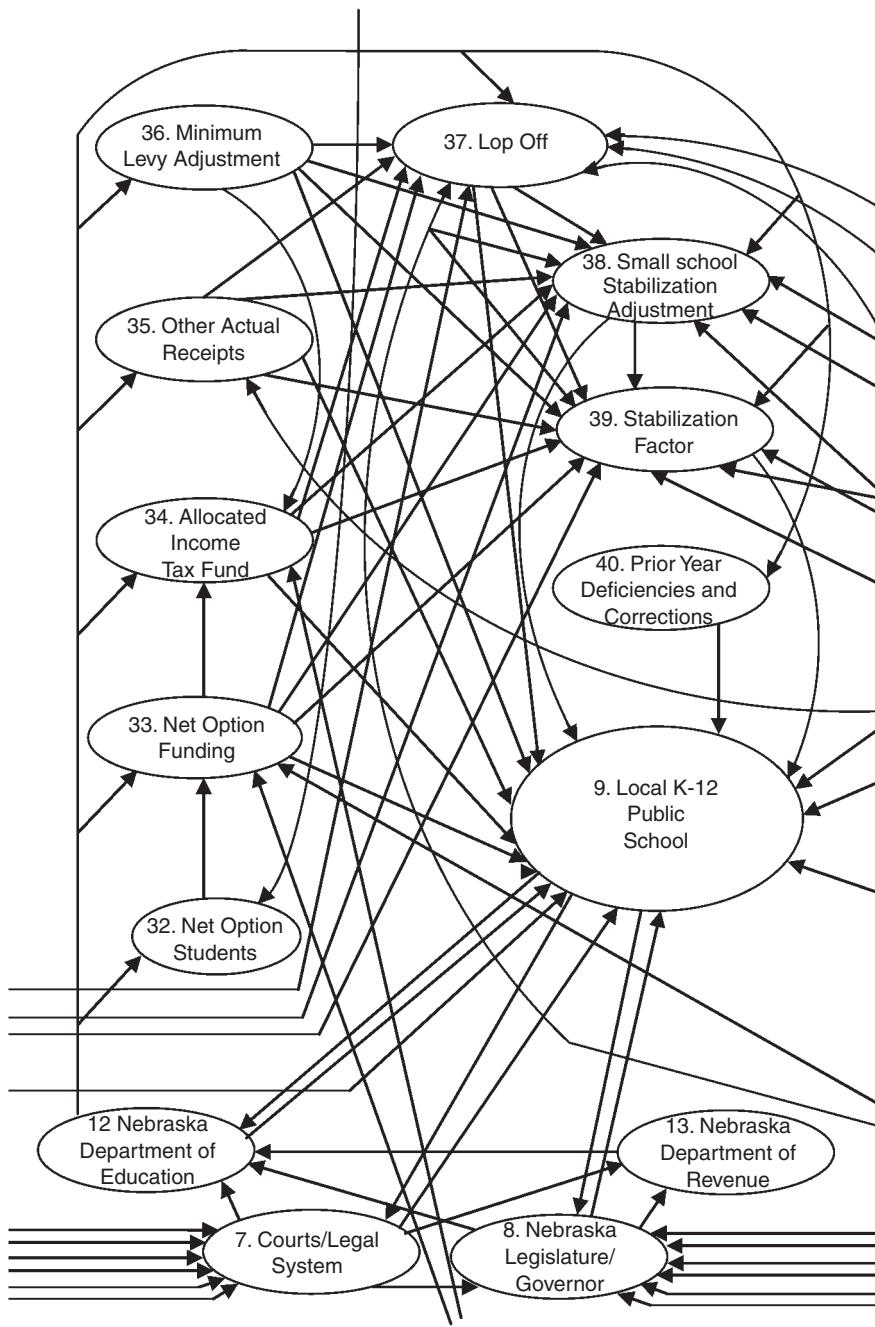


Figure 8.1 Social fabric matrix digraph components of Nebraska state aid for a local K-12 public school, 2006-7.

depicted in Figure 8.1, the term includes Net Option Students and deliveries from the environment outside the system. It calculates the amount of money received by a district because it has Net Option Students. The amount of money calculated is added to the local district's fund, as indicated in the delivery edge between components 33 and 9. This term is financially beneficial to wealthy suburban districts that receive large numbers of students whose parents can afford to transport children from their home district in the city. As outlined in Figure 8.1, the rules require that the Net Option Funding formula terms be incorporated into the Allocated Income Tax Fund, Lop Off, Small School Stabilization Adjustment, and the Stabilization Factor (Neb. Rev. Stat. § 79–1009).

- *Allocated Income Tax Fund:* The purpose of the Allocated Income Tax Fund is to distribute a special fund allocated from the state income tax among the 257 districts. The calculation includes formula terms from the system environment outside Figure 8.1 as well as the Net Option Funding and Minimum Levy Adjustment terms. Its findings are delivered to Lop Off, Small School Stabilization Adjustment, Stabilization Factor, and the local school districts. The Allocated Income Tax is both added to and subtracted from every school district; however, since the addition and subtraction take place in different sections of the formula, hold-harmless rules for some sections means the plus and minus of the same amount do not cancel each other out for all districts (Neb. Rev. Stat. §§ 79–1005.01 and 79–1005.02).
- *Other Actual Receipts:* The Other Actual Receipts term is a list of federal, local, and state receipts for the local school districts that are subtracted from what a local district is to receive from TEEOSA. The majority of Other Actual Receipts is the same list of receipts designated as Special Receipts and added to part of the formula outside the system in Figure 8.1. The plus of Special Receipts in the system environment and minus of Special Receipts in the system in Figure 8.1 cancel each other out and, thus, could be deleted from the overall formula without affecting district receipts. In addition to delivering its findings to indicate funding for local districts, Other Actual Receipts delivers its findings to Lop Off, Small School Stabilization Adjustment, and the Stabilization Factor to be used in their calculations (Neb. Rev. Stat. § 79–1018.01).
- *Minimum Levy Adjustment:* The set of rules making up the Minimum Levy Adjustment is applied to any local school district whose general fund property tax levy is less than 0.945. The calculated adjustment subtracts the local district's tax levy from 0.945 and multiplies the result by the adjusted property valuations divided by 100. If the Minimum Levy Adjustment is greater than or equal to the Allocated Income Tax Fund, the local district shall not receive monies from the Allocated Income Tax Fund. In addition, the Minimum Levy Adjustment organization delivers its calculations to Lop Off, Small School Stabilization Adjustment, and the Stabilization Factor (Neb. Rev. Stat. § 79–1008.02).



- *Lop Off*: Lop Off affects only districts that receive equalization aid. It is a set of calculations designed to limit state aid to some districts. The term is long and complex because of all the components explained above that feed into it and because of calculations delivered from outside this system that are included. Lop Off is designed so that no local system may receive equalization aid when total state aid, added to a levy of 0.95 times the prior year's adjusted property valuation divided by 100, exceeds numerous other calculated entities that are added together (for a list of entities, see Nebraska Department of Education 2006: 9). The Lop Off calculation is delivered to the local district, Small School Stabilization Adjustment, and the Stabilization Factor (Neb. Rev. Stat. § 79–1008.01).
- *Small School Stabilization Adjustment*: This component calculates aid distribution to school districts that have 900 or fewer formula students and adjusted operating expenditures per formula student less than the average for all such districts. The aid distribution is made proportionately to these districts based on a dollar difference calculated as follows: .8875 times the sum of the aid the district received in the preceding year plus property tax receipts for the preceding year, minus an amount equal to each district's calculated aid plus the product of 1.05 times the quotient of the district's assessed valuation divided by 100. The result of the calculation is delivered to the local district and the Stabilization Factor (Neb. Rev. Stat. § 79–1008.01).
- *Stabilization Factor*: This factor is calculated so that a district shall not receive state aid which is less than 83.75 percent of the amount of aid certified in the preceding year, minus an amount equal to 1.05 times any increase in the adjusted property valuation between the adjusted valuation used for the certification of aid in the preceding school fiscal year and the adjusted valuation used to calculate this year's state aid divided by 100 (Neb. Rev. Stat. § 79–1008.01).

## Findings for CCC/GSA

The empirical base provided by Figure 8.1 allows for conceptual findings as follows.

### *Three different patterns of CCC*

The description of Figure 8.1 identifies three different patterns of CCC. The first is accumulation through the system. Since the calculations are completed for each component there is a continuous cumulative effect, as those calculations are transferred to other components. This system characteristic is emphasized in macroeconomic studies, and is found in all systems. The second pattern is the set of inputs from the environment surrounding the open system, as indicated by the edges without specified components directed to components in Figure 8.1 from the environment outside the system.<sup>2</sup> All systems are open to an environment,

making all systems dynamic. The third pattern of CCC, which was Myrdal's main concern, is due to changes made to components inside and/or outside the system. He stated that circular causation implies that if one condition changes, "others will change in response, and those secondary changes in their turn cause new changes all around, and so forth" (1974: 730). For example, if local school districts successfully petition courts (edges 9 to 7 in Figure 8.1) to change a rule or set of rules provided by the legislative body for distributing state aid, flows can change both within the system and/or the system's environment. Thus, it is invalid to assume a single kind of CCC pattern.<sup>3</sup>

### ***Numerous deliveries to and from components***

Consistent with complex systems, there are numerous deliveries to and from components. Abstract models in the system dynamics literature often indicate only one delivery to and from each component in a system, often of an abstraction like a positive or negative sign. Such is not found in the real-world SFM in Figure 8.1. Abstract models in the literature about homeostatic control often have only one edge to each component without any specification about what is delivered among the institutional components. One delivery is assumed for coordination and control but that is not the real-world case. There are numerous different kinds of deliveries in and out of institutional organizations and often different sources for the same kind of delivery. The main purpose and activity of an institution is to serve other institutions. Thus, the viability of that service depends on what is being delivered, the level being delivered, and how well it fits with other deliveries. A positive or negative sign is not accepted as a valid and sufficient delivery in a real-world institutional setting, for example, in a school district.

### ***Individuals do not structure the rules or the system's environment***

Geoffrey Hodgson has demonstrated the continued popularity of ideas about rationality and individuals in the literature of economics (2007). Original institutional economists have documented that real-world cases are inconsistent with the model of rational individuals as traditionally attributed by economists, and the behavioral economics literature documents that people do not have mental capabilities to think that way. Humans are not rational (nor quasi-, semi- or demi-rational) and they are not individualized. Research conducted and observations made in the completion of the finance system in Figure 8.1 discovered neither rational calculators nor individuals. In fact, persons given agency positions by institutions to conduct themselves according to Nebraska rules and regulations would be in serious trouble if they attempted to act in such a manner. For personnel to complete their tasks, the standard operating procedures of the bureaucracy are utilized, not the preferences and wants of personnel. Individual means indivisible and separate, and neither characteristic is demonstrated by those in the school finance system. Rather than indivisible and separate, a person is divided and the parts are very integrated with other

system components and elements. An example is a person who is a professional accountant, consultant, active Catholic Altar Society member, wife, lobbyist, mother, family asset manager, Democrat, and Sierra Club member. For each activity she has different roles, particularly in the government department for which she consults. She is further divided by the activities she undertakes because of the overlay of systems such as the global oil cartel, climate change, and her country's wars. Divided and divided! Personal divisibility is reality.

Humans serve in assemblies of cognitive agents in order to complete the roles necessary for the collective behavior of socioeconomic processes. Thus, a more appropriate term for humans than individuals might be *homo nomos*. As Robert Cover stated: "We inhabit a *nomos* – a normative universe. We constantly create and maintain a world of right and wrong, of lawful and unlawful, of valid and void" (Cover 1984: 4). The actors found in the Nebraska school finance system act according to Cover's explanation, not according to that of Douglass North. North stated that "institutions are the external (to the mind) mechanisms individuals create to structure and order the environment" (North 1994: 363). That is not consistent with reality. CCC/GSA clarified, and the discussion about Figure 8.1 reconfirms, that workers and agents in the institutional components do not structure and order the environment outside the system. Neither do they order and structure the system itself nor its *nomos*. Those are created, delivered, regulated, and enforced by institutions – by the Nebraska Senate, courts, local school districts, NDE, and department divisions.

### ***Schumpeterian destroyer not found***

Findings from the school finance case can be utilized to comment on a disagreement regarding the emergence of new rules and systems.

Kurt Dopfer, John Foster and Jason Potts (DFP) stated: "Rules are the building-blocks of systems that form the micro-structure, or organization, of an economic system" (Dopfer *et al.* 2004: 267). This leads DFP to be interested "in the micro analysis of agents originating and adopting rules, and the complex structures that arise" (269) as a result. That process creates change because of the "interactions of agents as they adopt and use rules" (269). The agent is a Schumpeterian creator and, thereby, a destroyer of the initial order as the "agent originates, adopts, adapts, and reclaims a novel generic rule" (269). For DFP, this creative destruction and concomitant "economic evolution is a process of change in rules and rule-systems" (271). Its "*origination* is where an agent develops an idea/rule that leads to the design of an organization" (271). This is a case of "the Schumpeterian ... entrepreneur acting imaginatively" (273). "The first thing to happen in the micro domain is that an agent explores a new rule and its capabilities" (272). Consequently "we end up with a new micro organization" (272). The new evolutionary trajectory "begins when an agent acquires, understands and imagines how to use a novel rule" (272).

Wolfram Elsner expressed concern about theories based on an "*isolated Schumpeterian agent*" (Elsner 2007: 5). He explained that DFP assume that the

“originator of a new rule, however, appears to be some isolated Schumpeterian creative inventor” (2) and that the “creative destroyer has the ideas for the rule and he/she continually explores new ideas because the mind is restless” (5).

Elsner’s disagreement draws, in part, on the criticism of others who have explained that DFP fail to (1) take account of the fact that the design of and decisions about rules must be coordinated with a multiplicity of actors in a network of institutions located across overlapping systems; and (2) recognize that there will be different decision processes for rule innovation in different problem contexts (5). Elsner’s own concerns are: First, changing systems “involve *changes of the structure of incentives* to search, explore, experiment or imitate” (5) with which the agent would have to deal rather than remain isolated in a fixed setting. Second, the social rule not only has to be traced back to a defined complexity but also to uncertainty problems which have to be solved collectively (2). Third, institutions, as opposed to the adoption of rules by an entrepreneur in a micro context, “are used to solve coordination problems and thus carry new and jointly learned knowledge” (3).

The case of the Nebraska school finance system is inconsistent with the contentions of DFP. It is a system of overlapping institutions and organizations that are coordinated by laws, court decisions, regulations, monitoring, audits, media scrutiny, and so forth. Changing conditions are responded to by institutional procedures as new rules are designed, challenged, lobbied by adverse interests, tested, litigated, and adopted in a dynamic setting through the transactions of components of different systems. As found in the school aid case, explicit procedures and actions are taken to prevent any actor with an entrepreneurial inkling from changing rules that are codified by the whole process.

Recognition of CCC/GSA means the primary unit of analysis should be the components of the social fabric matrix. Agents (with agency power) are contained in institutional components. Too often, in economic literature, agents are defined and treated like individuals. An agent is not an individual. An agent is a person or party authorized and empowered by a principal to act in defined capacities. Agency is bestowed by the establishment of a relationship based on an understanding between persons and/or parties for an agent to act. Agents are not isolated or separate units because agency agreements are concerned with accountability, auditing, and performance. Most agents are institutional organizations (corporations, government departments, nonprofit organizations) that obtain agency status from other organizations. When persons are given agent responsibilities, they are located in an institutional organization and deal with a nomos.

DFP stated that “the agent has a mind” (2004: 269). First, an institutional organization such as a corporation does not have a mind. Second, persons involved have minds, not a mind.<sup>4</sup> Persons given responsibilities in an institutional organization transact with other persons in that organization and with those in other organizations, some of whom are located in other systems, so the agency function is a multi-system assembly of the knowledge, skills, and cognitive capabilities that are coordinated to perform the agency function. “Thinking is a social activity” (Menand 2001: 431).

## Conclusion

The Nebraska educational finance study illustrates how the SFM provides a methodology to utilize CCC/GSA and to further develop its principles. From such studies of real-world cases, we can continue to refine and develop socioeconomic concepts. This SFM application provides a refinement of our understanding of circular and cumulative causation patterns and of the definition of relationships among system components through deliveries. Furthermore, it confirms that the social prescriptions and proscriptions called rules are not the result of self-action by agents with fixed attributes.

## Notes

- 1 All statutes cited, along with the relevant list of source laws, may be found at the Nebraska Legislative Documents website (see References).
- 2 The environment to the system can be represented in a SFM as a single row that makes deliveries to the system's columnar components and a single column that receives deliveries from row components, or, components of the environment can be represented in separate rows and columns. In the latter case, to be consistent with CCC/GSA, the components in the environment would not receive from or deliver to other components in the environment. They would only make deliveries to or receive deliveries from system components.
- 3 It is also invalid to assume that cumulative causation is always circular. Some systems have delivered to surrounding environments for centuries without a return of circular impacts – soil erosion is an example.
- 4 What is often referred to as the human mind has so many different kinds of parts, divisions, and functions throughout the human body, and beyond, that it is probably not useful to designate that system as a singular closed concept or one that is separated from influences outside the human body.

## References

- Bailey, K. D. (1994) *Sociology and the New Systems Theory: Toward a Theoretical Synthesis*, New York: State University of New York Press.
- Cover, R. M. (1984) "Foreword: Nomos and Narrative," *Harvard Law Review*, 97: 4–68.
- Dopfer, K., Foster, J. and Potts, J. (2004) "Micro–meso–macro," *Journal of Evolutionary Economics*, 14: 263–79.
- Elsner, W. (2007) "Why Meso? On 'Aggregation' and 'Emergence' and Why and How the Meso Level is Essential in Social Economics," *Forum for Social Economics*, 36: 1–16.
- Hayden, F. G. (2006) *Policymaking for a Good Society: The Social Fabric Matrix Approach to Policy Analysis and Program Evolution*, New York: Springer.
- Hodgson, G. (2007) "The Revival of Veblenian Institutional Economics," *Journal of Economic Issues*, 41: 325–40.
- Menand, L. (2001) *The Metaphysical Club*, New York: Farrar, Straus & Giroux.
- Myrdal, G. (1944) *American Dilemma: The Negro and Modern Democracy*, New York: Harper & Row.
- (1974) "What is Development?," *Journal of Economic Issues*, 8: 729–36.
- (1978) "Institutional Economics," *Journal of Economic Issues*, 12: 771–83.

- Nebraska Department of Education (2006) *Tax Equity and Educational Opportunities Support Act Certification of 2006–2007 State Aid*, Lincoln: Education Support Services.
- Nebraska Legislative Documents. Revised Statutes. Chapter 79. Schools. Online. Available at: [http://uniweb.legislature.ne.gov/legaldocs/view.php?page=index\\_statutes](http://uniweb.legislature.ne.gov/legaldocs/view.php?page=index_statutes) (accessed July 2007).
- North, D. C. (1994) “Economic Performance through Time,” *American Economic Review*, 84: 359–67.

## 9 Unnatural depletion and artificial abundance

### A circular cumulative causation analysis of salmon fisheries and some implications for political ecological economics

*Sebastian Berger and James Edward Glavin IV*

#### Introduction

Since the second half of the twentieth century, in particular, the Western industrialized world has experienced an unprecedented economic “growth” that is customarily expressed in terms of GDP. Industrialization, manufacturing and growing markets have paved the way for exploiting the benefits of so-called “increasing returns” that are at the core of CC growth theories. In this very same economic process, however, the enormous wealth of biotic renewable resources, such as fisheries, has been reduced and exploited to an extent that has brought many of them near collapse and several species near extinction. Depletion of renewable and non-renewable resources as well as pollution are real costs that remain unaccounted for in markets. This is why an accounting in terms of market values underlying concepts such as “economic growth” (GDP in market prices) and “increasing returns” (in terms of exchange values) has been fundamentally criticized in the planning debate of the 1920s and 1930s, by the ecological development movement in the 1960s and 1970s, as well as by the movement for social cost accounting and green GDP in the 1980s and 1990s. In other words, the ecological context matters and economists must not turn a blind eye to the wealth-diminishing aspects of growth in manufacturing. In order to capture these important aspects, this chapter proposes Myrdal’s and Kapp’s CCC.

First, we identify common characteristics between Myrdal’s principle of circular cumulative causation (CCC) and the pre-analytical vision of ecological economics (EE), as exemplified by its treatment of fisheries. It is argued that there is a common “complexity perspective” that goes back to the 1940s when K. William Kapp applied the CCC to socio-ecological problems. Hereby, we demonstrate how CCC serves as a common denominator concept for institutional and EE. Second, we provide a circular cumulative causation analysis of the life cycle of salmon, and the social costs of salmon depletion and farmed salmon. Kapp’s CCC-based theory of social costs captures the wealth-diminishing



features of the economic system. The CCC perspective leads to a social minima approach in which states of social and ecological balance become part of economic rationality.

### **CCC as a methodology for fisheries**

Applying CCC in the ecological context raises the question as to whether it is compatible with EE's approach to complex renewable resources, such as salmon fisheries. EE's approach to fishery depletion argues that crafting institutional frameworks for sustainable resource utilization not only requires an understanding of the interactions between economic institutions and their effects on the ecological system, but knowledge of the complex interactions within the system of biotic resources itself.

Variables of an action situation [...] are also affected by attributes of the biophysical and material world being acted upon and transformed. What actions are physically possible, what outcomes can be produced [...] are affected by the world being acted upon in a situation.

(Ostrom 2005: 22)

"Appropriators in the field have to explore and discover the biophysical structure of a particular resource" (Ostrom 2005: 243). Thus, the attributes of resources themselves affect the sustainability of particular institutions so that the success of resource regimes depends on the characteristics of the resource. "With improved understanding, it may become possible to diagnose resource use situations well enough to separate promising institutional forms from those unlikely to achieve desired goals" (Dietz *et al.* 2002: 25).

Realizing the importance of knowledge about the resource situation, EE comes to adopt a "non-equilibrium" view of the ecological system in which fisheries are considered to be complex resources (Rose 2002: 241). This means that all relevant factors, such as habitat, spatial distribution of local stocks, population behavior, spawning stock biomass, the population size of other species, such as prey fish and predators and other ecological factors have to be considered (Wilson 2002: 331). Accordingly, EE holds that scientific research about causal relationships affecting the population size of fish bears many uncertainties, often preventing precise measurement and exact quantitative foundations for decision making (Wilson 2002: 329, 334–40). Fisheries are, furthermore, considered heterogeneous so that situational local knowledge is deemed important:

Fishers have to know a great deal about the ecology of their inshore region including spawning areas, nursery areas, the migration routes of different species, and seasonal patterns [...] inshore fisheries that have survived [...] have learned how to maintain these critical life-cycle processes with rules controlling technology, fishing locations, and fishing times.

(Ostrom 2005: 230)

Due to this complexity view, EE has serious doubts about the unproven “primum-mobile” assertion of formal resource economists that deems the spawning stock biomass the *only* relevant factor for sustaining stocks (Wilson 2002: 329, 331). EE rejects extremely simple, formal approaches of resource utilization, such as H. S. Gordon’s “equilibrium” resource model, and their conclusions. Gordon’s impact is still manifest in those approaches to fisheries that focus on the “maximum economic yield” of a single species, showing little concern for the complexity of the ecological system. Species are treated as isolated entities, essential reciprocal interrelationships are neglected, and fundamental uncertainty is reduced to stochastic uncertainty (Dietz *et al.* 2002: 9–10; Wilson 2002: 329).

Quite similar to EE’s complexity approach, Myrdal developed CCC to analyze complex socio-economic dynamics:

The dynamics of the ... system are determined by the fact that among all the endogenous conditions there is circular causation, implying that, if there is a change in one condition, others will change in response. [...] There is no basic factor; everything causes everything else. This implies interdependence within the whole ... process.

(Myrdal 1978: 774)

According to Myrdal, the task of CCC analysis is

to analyze the causal inter-relations within the system itself as it moves under the influence of outside pushes and pulls and the momentum of its own internal processes. [...] The outside forces push and pull the system continuously, and at the same time change the structure of forces within the system itself.

(Myrdal 1957: 18)

As a result, Myrdal argued that

it is useless to look for one predominant factor, a “basic factor” [...] everything is cause to everything else in an interlocking circular manner. [...] the application of this hypothesis moves any realistic study [...] far outside the boundaries of traditional economic theory.

(Myrdal 1957: 19)

This shows how Myrdal designed CCC as a framework of complex *causal* analysis in the Veblenian tradition rather than as a *teleological* search for levels of equilibrium or “primum-mobile” causation.

CCC accepts the time element and uncertainty as important characteristics of many complex dynamics: “because the system is in constant movement [...] the coefficients of interrelations among the various conditions in a circular causation are ordinarily not known with quantitative precision” (Myrdal 1978: 774). “The

time element is of paramount importance, as the effects [...] will be spread very differently along the time axis” (Myrdal 1957: 19).

Elements of inertia, time lags, and in extreme cases the total non-responsiveness of one or several conditions to changes in some set of other conditions are problems of great complexity [...] Consequently, our analysis [...] must often end in broad generalizations and merely plausible hypotheses, built upon limited observation, discernment and conjectural judgments.  
(Myrdal 1976: 83)

Myrdal defended the CCC’s methodological middle ground between the exact reasoning of mathematics and naïve empiricism against mainstream critics:

In calling the holistic approach the fundamental principle of institutional economics, I imply that our main accusation against conventional economists is that they work with “closed models” with too few variables. [...] institutional analysis, not working with the wholesale exclusion of so much which is important, will seldom be able to argue quantitatively in such precise terms, simply because our knowledge of relevant facts and the inter-relations between those facts is not that precise. [...] we are generally more critical than conventional economists. [...] We just want our theory and models, indeed even the concepts we use, to be more adequate to the reality we are studying.

(Myrdal 1976: 83–4)

Studying Veblen’s and Myrdal’s CCC approaches in the 1940s and 1950s, Kapp was probably the first economist to apply CCC to ecological problems (see his theory of social costs; Kapp 1950). In this context, Kapp applied the CCC as a working hypothesis, i.e. “a guiding principle of interpretation in the light of which we look at facts in order to see whether it improves our understanding,” broadening the frame of reference and also the unit of investigation (Kapp 1965b: 6). Kapp chose the CCC because it allowed the researcher to follow the lead of the subject matter, i.e. to view the socio-economic and ecological processes as the outcome of a circular cumulative process in which a number of causes interact upon each other and with their effects. “[Economics] is necessarily concerned with processes of considerable complexity. [...] Above all [CCC] directs our attention away from the futile search for the primary cause of [...] events” (Kapp 1965b: 7). His socio-ecological indicator approach to measuring advances in development perceived the quality of life as a totality in which everything is interlinked with the quality of the human environment. For example, pollutants in water and pollutants in air interact with one another in a reciprocal and even self-reinforcing manner to impact upon the quality of life in a cumulative manner. In addition, Kapp pointed to certain critical zones or thresholds in the ecological system, beyond which a change in a variable could have disproportionate and possibly irreversible effects (Kapp [1973] 1974: 100–1).

Kapp's CCC perspective is compatible with open system theory and the implications of the entropy law (2nd law of thermodynamics) as pointed out by Georgescu-Roegen (1971). They all view the economic system as embedded in the ecological and social system from which it receives its inputs and as affecting all interconnected systems. Importantly, CCC embodies the perspective that disruptive ecological effects can feed back and cause harm in the economic system (Berger and Elsner 2007). In conclusion, EE's approach to complex problems and fisheries and CCC's methodology are compatible. The CCC analysis of salmon in the following part of the paper is inspired by Kapp's proposal:

By viewing human action as taking place within, and with repercussions on a physical and social environment with specific structures and regularities, it becomes clear that the various spheres of man's environment which are affected by his action are interdependent [...] causal analysis cannot be carried on in terms of one or the other of the compartmentalized social, physical and biological disciplines [...] we still lack such a theory and/or science which is capable of elucidating the mode and outcome of the complex interaction of several systems. [...] However, there is one important aspect we do know about the causal chain [...] in many (if not in most) instances it is a process of circular causation which has a tendency of becoming cumulative unless some deliberate action is taken to arrest or redirect it.

(Kapp [1970b] 1983: 44)

### **A CCC analysis of salmon's life cycle, salmon depletion, and farmed salmon**

Applying the CCC to salmon's life cycle and salmon depletion, the following section demonstrates how the working hypothesis elucidates the complex reciprocal and self-reinforcing causation between salmon stocks and other species and plants in oceans, rivers, and land, as well as relations between ecosystems. Cumulative effects over time, such as the cumulative exhaustion of salmon stocks, and the likelihood of incomplete knowledge about interrelations are taken into consideration. The abundance of salmon stocks is explored not as an independent entity but with the hypothesis of being the result of circular cumulative causation, i.e. the cause and the effect of a balance in the ecological system exhibiting complexity and sensitivity. This analysis aims at providing an understanding of the ecological problem situation, setting the stage for a discussion of the social costs and underlying institutional causes of salmon depletion.

At alternate intervals of their life cycle, salmon utilize both marine and freshwater habitats, a life history strategy known as "diadromy." Born in lakes and streams, salmon migrate to the sea where they live for several years, then swim back upstream to spawn, and for many species to die.

- *Salmon's reciprocal interaction with the ocean system:* Young salmon arrive in the ocean after about a year of development in lakes, rivers, and streams.

At this point in the salmon life cycle, the new ocean dwellers, called smolts, already eat other fish, and as they grow they hunt prey that is larger and higher up the food chain. Smolts, however, are consumed *en masse* as they reach the ocean “by several species of marine birds and mammals, including seals, sea lions, and small whales [...] and some saltwater fishes, for example, walleye Pollock [...] and Pacific herring” (Willson and Halupka 1995: 493). Those that survive live on to accumulate over 90 percent of their adult body weight at sea, amassing a deposit of marine nutrients that proves incredibly important for inland terrestrial ecosystems. Salmon, fattened by ocean food webs, grow ripe with eggs, endure a final bulk-up and growth of secondary sex characteristics, and head inland, sometimes traveling hundreds of miles and thousands of feet in elevation to return to the streams in which they were spawned (Hildebrand *et al.* 2004: 1–2).

- *Salmon’s reciprocal interaction with the stream-forest systems:* As salmon travel upstream they are killed and eaten by a broad variety of terrestrial predators that rely on them as “a predictable, dependable, concentrated, and accessible resource high in protein and energy” (Hildebrand *et al.* 2004: 2). Moreover, Hildebrand *et al.* argue that

spawning salmon has also been important in supporting the nutrient requirements, particularly nitrogen, of periphyton, juvenile salmon, and resident fishes. Growth rates of juvenile fish in streams containing spawning coho salmon (*Oncorhynchus kisutch*) were double those that lacked returning fish, and the proportion of salmon-contributed nitrogen in the tissues of freshwater biota ranged from 17 to 30% across trophic levels.

(*ibid.*: 2)

Those salmon that die as a result of the rigors of upstream travel, or are killed and incompletely consumed both by their predators and the scavengers that follow, end up depositing their nutrients into forest flora. The impact of this deposition is extensive, and it has been observed that almost a quarter of the nitrogen embodied in the trees and shrubs within 150 meters of salmon-running streams come from marine sources. That is, a quarter of the nitrogen in some forests is left by salmon. As a result, vegetation within range of this marine nitrogen deposition is shown to achieve triple the growth rate at spawning sites relative to reference sites (Helfield and Naiman 2001: 2406). Shading by riparian vegetation moderates stream temperatures, controlling rates of embryo development. Bank stabilization by riparian roots minimizes erosion, which threatens embryo survival by restricting intragravel flow and oxygenation of redds. Litter inputs provide allochthonous organic matter supporting production of aquatic insects, which are an essential food source for juvenile salmon. Riparian forests also enhance stream habitat through the production of LWD (large woody debris). Among other functions, LWD increases the structural complexity of stream channels, thereby creating preferred habitat for spawning. Instream LWD also creates areas of low flow velocity, providing shelter from winter high flows, which is an important cause of mortality in overwintering

fry and incubating embryos. Overall, the presence of LWD in spawning streams enhances production of salmonids fishes (Helfield and Naiman 2004: 2407).

Consequently, Helfield and Naiman argue that enhancing the growth of riparian trees and the production of LWD inputs to the riparian zone serves as a self-reinforcing causation in which spawning salmon help to enhance the survivorship of subsequent salmonid generations, as well as the success of terrestrial and aquatic predator species and proximal flora. It is apparent that such a system operates cumulatively; that is, salmon runs in one year affect the next, and salmon runs in that year affect runs far into the future.

This analysis of the salmon life cycle, and its cooperative integration with the surrounding ecosystems, provides the basis for illustrating the disruptive effects of unsustainable industrialization as evidenced in case studies from Maine, the Clackamas River Basin in Oregon, and the Rhine River in the Netherlands. Salmon is a valuable economic resource in the modern human economy because it is one of the largest commercial fish and one of the healthiest sources of protein and fat in the human diet (Myers and Worm 2003: 280). Fish resources, in general, are of paramount importance to the human economy because three billion people around the world currently rely on fish protein for at least 20 percent of the animal protein in their diet (Mercer 2006: 5). However, only a few regions still produce wild salmon abundantly and some scientists even project that by 2050 there may be no commercially viable fisheries left anywhere in the world (Worm *et al.* 2006: 790).

As one of the largest salmon tributary systems in the United States, the Clackamas River Basin terrain was heavily timbered and rocky, and development in the basin was slow until the late 1800s when industrial machinery cleared the way for settlement. Salmon populations were damaged by timber harvest, road building, driving logs downstream, dam building, removal of riparian vegetation, and the destruction of side channels and wetlands. These changes contributed to dramatic losses of salmon habitat and productivity:

the annual return of ... salmon and trout to the Columbia River Basin has decreased from an estimated 12–16 million individuals in the 1880s to 2.5 million in the 1980s (NPPC 1987). Furthermore, Nehlson *et al.* (1991) identified 214 stocks of Pacific salmonids from California, Oregon, Idaho, and Washington that they considered to be of special concern, as they face a high or moderate risk of extinction.

(Hildebrand *et al.* 2004: 5)

Maine has the oldest and last Atlantic salmon fishery in North America.

Historically, 34 rivers and streams in Maine had naturally reproducing Atlantic salmon populations. [...] Before the construction of dams in the early 1800s, the upstream migrations of salmon extended well into the headwaters of large rivers [...] vast amounts of habitat in the headwaters of the Penobscot and Kennebec Rivers were used for spawning and rearing.

(Saunders *et al.* 2006: 540)

Organized fishing was accomplished mostly with the use of weirs, which contributed to the relatively rapid depletion of the fishery. By the late 1800s output peaked, and in 1948 only 40 salmon were caught. Maine's "catch-and-release only" system was shut down in 1999 for fear of endangering the very limited salmon population (Kocik and Sheehan 2006: 3).

The Rhine River in the Netherlands has hosted salmon species probably since the retreat of the glaciers which carved the river itself. Modern fishing efforts for salmon in the Rhine were typical, steady progressions until "the end of the 1920s witnessed a major decline in salmon catches," and eventually, "the Dutch salmon fishing industry effectively ceased to exist after 1933" (De Groot 2002: 213). Massive restocking programs since the 1930s have not proven successful, suggesting a failure to comprehend the particular needs of salmon – for example, "[it] is not clear whether certain smells that are characteristic of the spawning grounds are still discernible to salmon that have passed through rivers rich in detergents" (De Groot 2002: 215). Much of the continued depression of salmon abundance is also caused by a large number of locks and weirs, canalization and flood control, sand and gravel extraction, and wastewater discharges (De Groot 2002: 215).

The above case studies identify organized overexploitation of salmon and the elimination of habitat through construction as the combined causes of a cumulative disruption of the salmon life cycle and salmon depletion. As early as 1950 Kapp's understanding of CCC led him to detect the importance of maintaining states of complex ecological balance and guaranteeing sustainable rates of exploitation: "these far-reaching and often irreversible consequences of human activities are all due to the fact that they disturb the complex and highly sensitive ecological balance" (Kapp [1950] [1963] 1977: 95). In particular, Kapp observed the vulnerability of salmon fisheries due to their breeding habits and the possibility of cumulatively diminishing stocks: "whenever the rate of utilization exceeds the critical limit [...] the resource flow tends to decrease and the resource may finally lose its capacity to renew itself altogether" (Kapp [1950] [1963] 1977: 93).

The decline of wild salmon catches was followed by farmed salmon so that "in only a short time, salmon have gone from abundance to depleted stocks to abundance again. What has made this possible is aquaculture" (Clay 2004: 515). "World salmon and salmon trout supply increased more than four-fold between 1980 and 2001 and the share of North American wild salmon in total world supply fell from more than half to about one-sixth" (Knapp *et al.* 2007: xv). Today, over half of the salmon consumed in the world is farmed. Farming salmon requires huge and constant inputs of both non-renewable sources of energy and materials, and releases wastes concomitant with the scale of inputs. In their infancy, salmon are kept in heated tanks to encourage growth. Soon after they grow into smolts, they are individually vaccinated (Clay 2004: 516–18). Research into farmed animals has observed that agricultural and veterinary medicines have cumulative effects in ecosystems as they are excreted out of target animals and into local watersheds (Daughton and Ternes 1999: 921).



When the salmon finish their yolk sacs, they are fed protein pellets made from “junk fish,” which are commercially unviable fish species, or by-catch. This diversion of food from ocean ecosystems to human enterprise not only has a negative impact on ocean food webs, but

the input of fish products is two to four times the volume of fish outputs for these crops. Because of their dependence on wild-caught fish and shrimp, salmon aquaculture deplete rather than augment fisheries resources [...] about 1.8 million tons of wild fish for feed were required to produce 644,000 metric tons of Atlantic salmon – a 2.8: 1 ratio [...] Consequently, [salmon farming] depends heavily on fishmeal imported from South America.

(Naylor *et al.* 1998: 883)

As the salmon grow to maturity they are moved from freshwater ponds to salt-water pens on coasts. Here they are kept at high densities, where “water circulation [...] washes away feces, uneaten feed, and other wastes. At present, waste disposal costs farmers nothing” (Clay 2004: 519). While the farmers may not suffer the effects of the salmon’s effluent, ocean ecosystems have long been known to be susceptible to artificial nutrient inputs, as in the case of nitrification and the Gulf of Mexico’s “dead zones.” During the adult lifespan of these salmon, cage failures and escape are not uncommon; in fact, farmed Atlantic salmon have been found as far as the South Pacific (Naylor *et al.* 1998: 883). These escapees also result in “interaction of farm with wild salmon [which] results in lowered fitness, with repeated escapes causing cumulative fitness depression and potentially an extinction vortex in vulnerable populations” (McGinnity *et al.* 2003: 2443). Finally, farmed salmon may present a significant threat to health. Salmon raised in captivity have never been studied to explore the effects of their close-proximity lifestyle and their ability to cope with continuous exposure to concentrated feces throughout their life cycle. There is reason to suspect that the diseases these salmon suffer as a result of aquaculture (Naylor *et al.* 1998: 883) have effects on their final consumers. A slab of farmed salmon also represents inputs of untested veterinary medicines, and there are “health risks (based on a quantitative cancer risk assessment) associated with consumption of farmed salmon contaminated with PCBs, toxaphene, and dieldrin [which] were higher than risks associated with exposure to the same contaminants in wild salmon” (Foran *et al.* 2005: 552). The cumulative and synergetic effects of different pollutants in animals and humans are difficult to determine with exactitude but can lead to great social costs in the form of adverse health effects. As yet another example of biological cumulative effects, salmon are exposed to these chemical compounds through their natural predation, wherein top-level predators like salmon ingest contaminated prey whose life-history strategy is essentially to collect toxins through its own feeding.

The analysis of salmon depletion and subsequent farmed salmon abundance illustrates how the disruption of ecological balance is cumulative and directional, heading towards greater ecological degradation. The CCC analysis of salmon

depletion shows a process in which certain economic organizing principles and modern industrial fishing techniques in circular causation disrupt the complex ecological balance of the salmon life cycle, causing salmon depletion with complex negative effects on interconnected species and plants and ecosystems. The cumulative causation of economic organizing principles and technology triggers a further downward spiral of ecological degradation via farmed salmon. In the latter the interaction between pollutants is synergetic and cumulative, there are time lags during which they accumulate in fish, humans and the environment, and there is uncertainty about the exact nature of their interrelations as well as the magnitudes of hidden effects. The heavy dependency on non-renewable energy inputs makes the production process of farmed salmon a wasteful and inefficient enterprise, certainly when compared to wild salmon. The protein inputs are imported from already relatively poor regions, further jeopardizing ecological balance in these regions and diminishing their potential to satisfy their own basic needs.<sup>1</sup> Moreover, the pollution of the oceans through aquacultures is an instance of ecological degradation at a rate that is higher than the natural rate of restoration and possibly irreversible.

### **The vicious circle of salmon fisheries: its social costs and institutional roots**

The CCC perspective may be considered one of the main reasons why Kapp adopted a definition of social costs that is broad enough to cover not only economic costs that may be expressed in monetary terms but also effects of less tangible character, such as effects on the impairment of human health and the quality of life: “the term ‘social costs’ covers all direct and indirect losses sustained by third persons or the general public as a result of unrestrained economic activities” (Kapp [1950] [1963] 1977: 95). Due to the co-existence of the private and public sector social costs are harmful effects of private and public economic decision-making (Kapp [1965a] 1983: 10). Kapp argued that “the decline of [...] salmon fisheries [...] is another example of substantial social losses [i.e. social costs]” (Kapp [1950] [1963] 1977: 140).

“The consequences will be not only a rapid and cumulative exhaustion but an irreversible process which may have far reaching adverse effects for the prosperity of entire industries and communities” (Kapp [1950] [1963] 1977: 93). One of the adverse effects, for example, is that alternatives of economic activities open to a social group are eliminated. Enforced specialization in economic development and arrested growth are serious consequences and constitute substantial social losses. “Thus communities may be left stranded and ghost towns may take the place of formerly striving settlements” (Kapp [1950] [1963] 1977: 93). And even where restoration is technically feasible, salmon depletion can be economically irreversible owing to its prohibitively high costs (Kapp [1950] [1963] 1977: 94). This process of social costs is tipped off by resource depletion and cumulatively leads to societal crisis and economic stagnation. This is precisely what Myrdal’s CCC analysis described as a vicious circle or “social waste,” leading to growing regional economic inequalities and poverty<sup>2</sup> (Myrdal 1957: ch. 3).

However, salmon depletion has led to a relatively new “social cost-phenomenon,” i.e. aquacultures and farmed salmon. When Kapp described disruptive farming practices, he could just as easily have meant farmed salmon:

these operations may have harmful effects which often fail to be considered by those interested in increasing [...] [the yield of salmon]. Indeed, if carried out on a large scale and without protective measures in response to a rapidly growing demand for [...] [salmon], the process of bringing [...] [farmed salmon] into use may endanger the very prosperity which it seeks to promote.

(Kapp [1950] [1963] 1977: 94)

The artificial abundance of salmon stocks produced by means of salmon farming emerges as a result of unnatural depletion of wild salmon so that both become cumulative causes of significant social costs that are ultimately born by society at large, third parties and future generations. The salmon analysis supports Kapp’s tenet that competitive conditions fail to guarantee that the harvest of salmon will not be intensified beyond a safe minimum rate. They also lead to more devastating practices aimed at making up for declining returns:

Not even [...] lower returns resulting from excessive production will necessarily lead to a curtailment of output. Quite the contrary, lower yields may actually prove an incentive to greater efforts [...], inducing them [fishing industry] to make up for their declining return by still greater output.

(Kapp [1950] [1963] 1977: 111)

The price system as such can be considered incapable of rationally defining an ideal output position regarding renewable resources, and has a devastating tendency to accelerate the rate of use of resources. In fact, reliance on the price system may lead to maximizing future costs and minimizing future benefits (Kapp [1950] [1963] 1977: 113).

Evidence of the problematic relationship between the “free” price system and the degradation of renewable resources that enter the process essentially as free gifts of nature has been found by Marx long ago, in particular with regard to price fluctuations: “the whole spirit [...] toward the immediate gain of money contradicts agriculture [and fisheries], which has to minister to the entire range of permanent necessities of life required by a network of human generations” (Karl Marx, *Capital*, Vol. III; Kapp [1950] [1963] 1977: 101). Likewise, Veblen recognized that the “American plan or policy is a settled practice of converting all public wealth to private gain on a plan of legalized seizure” (Veblen, *Absentee Ownership*; Kapp [1950] [1963] 1977: 139). “Business enterprise has run through that range of natural resources, the fur-bearing animals [and salmon], with exemplary thoroughness and expedition and has left the place of it bare. It is a [...] concluded chapter of American business enterprise” (ibid.: 137).

In the tradition of Veblen's *Theory of Business Enterprise* (1904), Kapp's institutional analysis identifies the situation of the modern (food) corporation as the main underlying cause of social costs. That is the application of modern technology under the guidance of pecuniary drives that are firmly entrenched in the system of business enterprise, coupled with a lack of effective public controls to safeguard ecological and social balance. According to this perspective, social costs are neither accidental, nor minor side-effects that can be fixed by ad hoc measures. Instead, they are systemic effects caused by pecuniary exigencies. The latter result from the capitalization of productive equipment and putative future earnings as collateral for monetary debt contracts (capitalization is based on an accounting system in terms of exchange value). The growth-logic of the interest rate requirement of debt contracts means that the corporation needs to generate a permanent and increasing flow of financial returns and is thus forced to increase its earning capacity. This requires a minimization of idleness in the expensive production process operating with extremely productive modern technology on a large scale (economies of scale) (Kapp, unpublished). In this perspective the corporation's drive for increasing pecuniary returns is institutionalized in the competitive system via a self-reinforcing causation of large-scale modern technological competition and monetary debt relations.

Applied to biotic resources, this growth rationale causes depletion because the sustainable harvest rate is relatively stable. Georgescu-Roegen's production theory explains why renewable biotic "funds" exhibit a specific service rate that does not grow exponentially (Georgescu-Roegen 1971). The pressure of the time value of money in property-based market regimes encourages depleting renewable resources that grow at a rate below the market interest rate (Steppacher forthcoming; Swaney 2006: 116) Salmon farming may be seen as an attempt to encourage biotic resources to conform to these technological and pecuniary exigencies, making fishing results predictable at a growing flow rate. This can, however, only work to the advantage of the individual firm when it is legal to ignore negative ecological effects and to legally shift these costs to society at large. The costs of depleting natural salmon, pollution and ecological degradation are unpaid for by the business enterprise, i.e. they are socialized. The now cheaply available farmed salmon further contributes to the creation of irrationally high consumer demand for salmon via sales promotion. In fact, also this kind of ecologically harmful consumption may be considered the result of technological and pecuniary exigencies of the corporation. The latter is forced to find ever new outlets for its enormous productive capacity as well as new sources of pecuniary profits (Kapp [1950] [1963] 1977; Steppacher forthcoming). Kapp's social costs theory points out that shifting costs to society where no charges for waste and depletion of food webs are levied is often a precondition for maintaining the earning capacity of the business enterprise. In such a system, pecuniary profits are made via wasteful production and the promotion of wasteful consumption while society's capacity to provide for its social reproduction and that of future generations is diminished. This peculiar situation was first pointed out

by Veblen, who defined waste not by means of a moral standard but as those practices that “do not serve human life or well-being on the whole” (Kapp [1950] [1963] 1977; Veblen [1899] 1917: 97). In this case profits and waste are not a contradiction but, in fact, condition one another.

### **CCC’s “real term” method and its implications for the theory of social costs**

Kapp’s findings lead him to develop a theory of social costs that stresses “the cumulative character and complexity of the causal sequence which gives rise to environmental disruption and social costs” (Kapp 1970a: 837) and “the delicate system of interrelationships” (Kapp [1950] [1963] 1977: 94). His theory can in this sense be coined a CCC theory or complexity theory of social costs that is able to deal with less tangible yet important effects and interrelationships (Özveren 2007: 203–4). Kapp emphasizes that social costs cannot be considered as minor side-effects in relatively isolated locations, but have to be seen as all-pervasive effects. He even formulates a historical hypothesis: “as the economy becomes more complex, non-market interdependencies are likely to assume greater significance. For this reason social costs are bound to become increasingly important” (Kapp [1965a] 1983: 5).

Kapp’s theory of social costs implies a holistic view of the quality of life (i.e. of the whole situation of private and social costs and benefits), and the dependence of societal reproduction on states of socio-ecological balance. It may also be called “substantive” in the sense that it aims at revealing negative effects on society’s reproductive capacity and on the satisfaction of fundamental human needs in real terms (Berger 2008b). This is in the tradition of Myrdal’s CCC-based insight that only a real term analysis as opposed to a monetary analysis can capture social costs and the condition of the whole situation (Myrdal 1957). Any deficiencies in this regard are a substantive indicator of social costs in real terms. Kapp’s and Myrdal’s argument for a substantive accounting of social costs is in the broad tradition of classical and institutional economics and their approach to surplus and distribution (Kapp, unpublished).

Kapp’s rejection of utilitarian and formal solutions to social costs (i.e. his critique of the Pigouvian and Coasian frameworks) is partly the consequence of the complex CCC perspective. The CCC perspective gives rise to a specific conceptualization of the problem of social costs that is much broader and follows the subject matter more closely and stays closer to the real facts. “An element of inescapable indeterminacy may remain either due to the lack of homogeneity of the facts or of people’s valuations or due to a lack of knowledge about causal interrelationships” (Kapp [1972] 1977: 309). Kapp pointed out why in particular circular cumulative causation in the ecological system and fundamental uncertainty render formal approaches inappropriate.

Pollution and the disruption of the environment are the results of a complex interaction of the economic system with physical and biological systems

which have their own specific regularities. Moreover, pollutants from different sources act upon one another and what counts are not only the effects of particular effluents and toxic materials but the total toxological situation. [...] Those who have studied these complex relationships know that environmental disruption can easily become cumulative with pervasive and disproportionate effects per unit of additional pollutants.

(Kapp [1972] 1977: 314–15)

Under such conditions the price system cannot identify a viable output position because complexity and uncertainty make it largely impossible for “the individual to ascertain the full range of short and long run benefits of environmental improvements, or for that matter, of the full impact of environmental disruption upon his health and well-being” (Kapp [1972] 1977: 314). On the other hand, leaving out less tangible social costs because they cannot be priced would make economics an incomplete, i.e. an irrational system of thought. Referring to Myrdal, Kapp reminded us “that statistical convenience and measurement must *not* be permitted to set limits to concept formations and thus to exclude relevant elements [...] to define the concept more precisely than is justified is logically faulty” (Kapp [1972] 1977: 309–10).

Kapp rejected the application of the monetary calculus that is based on the utilitarian principle (i.e. the willingness to pay or accept compensation) that assigns monetary exchange values to ecological effects. This method is, according to Kapp, as arbitrary as the distributional inequality which it expresses because the “willingness to pay” depends on the “ability to pay” and has nothing to do with the objective and real exigencies of states of socio-ecological balance.

The use of the willingness to pay as criterion of quantifying and evaluating the quality of the environment has the insidious effect of reinterpreting original human needs and [socio-ecological] requirements into a desire for money and of evaluating the relative importance of such needs in terms of criteria which reflect the existing inequalities and distortions in the price, wage and income structure.

(Kapp [1972] 1977: 313)

Kapp endeavored to show that monetary criteria, such as the willingness to pay, are not appropriate because they do not evaluate the characteristics which define the quality of the environment and its potentially negative impact on human health, human well-being, and human survival (Kapp [1972] 1977: 316).

Neither Coase nor Pigou aim to guarantee the fulfillment of the requirements of socio-ecological balance or the satisfaction of basic human needs but prescribe a fixed behavioral rule that is assumed to solve the problem. Their solutions are predetermined by their formal apparatus and apply only ex-post. This “catching-up” with the effects ex-post can be too late if damaging effects are irreversible, and can be too little as their closed system methodology prevents them from taking into account the whole range and full extent of repercussions:

making the content and extent of the control of environmental quality dependent upon individual willingness to pay could at best lead to piecemeal measures and an ineffective formal sub-optimization if it does not become the pretext for endless delays or a policy of doing too little too late.

(Kapp [1972] 1977: 315)

### **From the social costs of increasing pecuniary returns to sustainable socio-economic and ecological returns: the quest for new forms of democratic governance**

Social costs raise questions regarding the Kaldorian current of CC theory: how far is the promotion of increasing pecuniary returns sustainable and at what point does the system actually exhibit greater total costs than total benefits (total including the sum of social and private)? In other words, where are the limits of increasing pecuniary returns? Interestingly, Kaldor's theory of increasing returns notes that "agriculture (and mining) [and fisheries] produces both direct and indirect inputs for industry – basic materials and food. If agriculture is subject to the Law of Diminishing Returns, agricultural output may be constrained by land and the available technology" (Kaldor 1974 [1978]: 206). However, Kaldor also stated ambiguously that "total output can never be confined by resources" (Kaldor 1972 [1978]: 194) while "a maximum rate of growth [...] must be on account of the scarcity of natural resources, and the impossibility of substituting capital goods for natural resources at more than a certain speed" (Kaldor 1972 [1978]: 195). Based on his distinction between manufacturing industry (increasing returns) and agriculture (diminishing returns), Kaldor elaborates the consequences for development, competition, price formation, and trade (Kaldor 1978: xxii–xxiii). Yet, unlike Kapp, he does not analyze the social costs that result from a manufacturing system that attempts to generate increasing pecuniary returns by harvesting funds of renewable resources at a rate that is faster than their rate of reproduction. In addition, the fundamental difference between manufacturing and agriculture as well as the limits this imposes for economic growth were elaborated more accurately by Georgescu-Roegen (Georgescu-Roegen 1965 [1976], 1970). Thus, CC theories of increasing returns need to be embedded in a wider ecological and institutional perspective in the tradition of Kapp and Georgescu-Roegen.

As a solution to the problem of social costs, Kapp proposes safe minimum rates for the utilization of resources and maximum tolerance levels for pollution. He considers the latter to be social minima because they help avert the severe social consequences of resource depletion and pollution, such as arrested economic development and structural depression of entire regions.<sup>3</sup> The social minimum standard serves as a bench-mark for policy makers to evaluate private practices and public policies. It also serves to monitor the effects of different institutional and technological set-ups (Kapp [1950] [1963] 1977: 114). Kapp proposes the adoption of more selective practices of resource use that are in harmony with the life and growth cycle of salmon fisheries:



A wide variety of measures are available ranging from educational persuasion, closed seasons for [...] [fishing], specific interdictions, the subsidization of substitutes in plentiful supply, taxation, price control, rationing and outright prohibitions placed on the use of certain materials for low priority uses [i.e., luxury].

(Kapp [1950] [1963] 1977: 154)

The determination of social minima and the means to achieve them takes place in democratic political processes with the help of the sciences and economics. Kapp describes this as a process of democratic political economics (Berger and Forstater 2007; Berger 2008a) which contributes important political elements to EE (Özveren 2007: 190). Political economics also addresses the problems and exigencies that power asymmetries and vested interests inherent in the economic system pose for this democratic process (see Kapp's application of François Perroux' theory of the "domination effect" (Perroux 1950); Kapp 1968).

### ***Prospects for an institutional-political EE approach to fisheries***

While many ecological economists consider Kapp an exceptional economist whose very early emphasis on basic uncertainty and interdependencies of social and environmental systems is foundational to EE, the relatively weak influence of this branch of institutional economics (Röpke 2005: 278–9) has led to an integration based mainly on the neoclassical branch of new institutional economics. CCC's contribution and potential regarding the importance of "interdependence" in relation to environmental problems seems to be largely overlooked (see e.g. Paavola and Adger 2005: 354–7). Despite the fact that institutional economics in the US has not had an unproblematic relationship with EE (Swaney 1985: 854; Swaney 2006: 113, 122 n. 12)<sup>4</sup> both are increasingly being integrated (Barnes 2006). In Europe, on the other hand, institutional economics in the tradition of Kapp and Georgescu-Roegen was synthesized with ecological political economy in the tradition of Marx and Georgescu-Roegen (Guha and Martinez-Alier 1997, Steppacher forthcoming). In fact, Kapp had identified the common concerns of Marx and Veblen about ecological degradation early on. In particular, parts of Volume III of Marx's *Capital*, which deal with soil degradation, share similarities with Georgescu-Roegen's work on viable agriculture (Kapp [1950] [1963] 1977: 101; Foster 2000; Georgescu-Roegen 1965). Several European contributions also adopt Karl Polanyi's concepts of "embeddedness," "fictitious commodities," and "substantive economics" as a basis for integrating institutional economics and EE (Barthelemy and Nieddu 2007; Adaman *et al.* 2003). Similarities between Polanyi's and Kapp's concepts exist (Swaney and Evers 1989; Berger 2008b) and are considered to have potential for a further integration of institutional economics and EE (Özveren 2007: 191).

Looking at how EE explores sustainable institutional arrangements for resource utilization shows the close links to Kapp's work. EE applies economic policy analysis to empirically elaborate institutional settings that work to

conserve the resources (Ostrom 2006: 760). EE evaluates institutional performance according to multiple criteria including efficiency, sustainability, and equity. According to Dietz *et al.*, economic efficiency focuses on the total relationship of individual and social benefits versus individual and social costs, and does not presuppose exact quantitative measurement (Dietz *et al.* 2002: 25). The focus on social costs and social minima is considered to be a result of the complexity perspective (Stern *et al.* 2002: 463). In all of this, the goal of sustainability and livelihood in general prevails over individual economic profit making (Tietenberg 2002: 199–200; Plummer and Armitage 2007: 69).

Nevertheless, the question remains how to scientifically elaborate social minima in complex systems under conditions of fundamental uncertainty. The scientific determination of social minima is complex because much information is needed that often relies on experience, knowledge, and participation of the users themselves. This process works best when all participants take an interest in an integrated approach towards resource preservation that considers interspecies relationships and ecological balance. In general, research into complex systems finds that different problem contexts require different caps (social minima) and that intense use should be followed by an early shift away from the resource at signs of trouble, allowing the resource to recover (Rose 2002: 241). With varying ecological conditions there is also no single institutional blueprint and the idea that there is only one best solution is a great obstacle against creativity when the development of regimes is at stake. “This is the case for those saying that private ownership and market allocation is always the best solution, as it is for those believing that the state can allocate all goods or that more community is the solution to any problem” (Vatn 2005: 417–18). Consequently, EE proposes a diversity of methods to achieve collaborative environmental management.

### ***Free market environmentalism and EE’s critique***

Contrary to EE, free market environmentalism (FME) which is a branch of neo-classical environmental economics proposes one single “cure-all” approach, i.e. individual tradable quota regimes (ITQs). ITQs are individual property rights and are supposed to lead to sustainable practices due to the self-interested actions of the owners. The argument is that “Command and Control” regimes run by the state are “economically inefficient.” The main focus of FME is – as is not surprising – directed towards increasing the pecuniary profitability of the resources. FME proponents display almost no concern for the actual compatibility of this regime with the specificities of the resources and ecological balance. Its beneficial character is simply taken for granted or assumed: perhaps as an example of ecological “invisible hands.” From the CCC perspective, the main problem of ITQs is that they tend to neglect a long-run focus and complexity. Simple individual tradable property rights that can manage complex resources efficiently are hard to design because externalities are interlinked, so that complex institutional settings are needed for complex resources and complex user communities (Stern

*et al.* 2002: 463). Overall, FME seems to ignore the implications of CCC. However, even FME presumes a total allowable catch, i.e. the safe ceiling that can be harvested (cf. also Adaman *et al.* 2003: 363; Swaney 2006: 119). Existing ITQ regimes set total caps based, e.g. on historic practices and negotiations (Rose 2002: 236, 241). Unfortunately, FME proponents, such as Leal, do not explain how the total allowable catch is determined and do not address the apparent theoretical inconsistency (Leal 2005: 6). Social minima constitute a break with the market mechanism because they are social and community rights, and the government is more involved than in rights for commodities and individual property rights. The political process has to determine a schedule that specifies ceilings for each year going forward, and decides how they are allocated/distributed, also considering the possibility of technological controls to observe the ceiling (Swaney 2006: 118, 120).

### ***From “command and control” to “community-based management” regimes***

A case study from Alaska which hosts one of the few remaining sources of wild Pacific salmon is an example of a “command and control” regime. Alaska’s salmon resources as a whole are still relatively fecund. In the past 25 years Alaskan wild salmon harvests have remained high, and have even enjoyed a slight increase. This maintained productivity may be attributed to Alaska’s tight regulation of its 26 fisheries, using a “limited entry” program,

which was established in the 1970s to limit growth in the number of people fishing in the salmon industry. [...] In each fishery, fishermen may use only the type of gear specified by the permit. There are also numerous other restrictions on boats and gear. For example, in Alaska’s Bristol Bay drift gillnet fishery boats may not exceed 32 feet in length and gillnets may not exceed 150 fathoms (900 feet) in length. Individuals may hold more than one salmon permit, but they may participate in only one salmon fishery per season. A boat may only be used in one salmon fishery per season.

(Knapp *et al.* 2007: vi)

The future of Alaska’s salmon fishery is however still subject to uncertainties, because like all other fisheries, sustained yield is not a reliable indicator of a sustainable fishery and in fact, most fishery collapses exhibit high or moderate production until their crash rather than a slow decline (Mullon *et al.* 2005: 111).

Despite its fundamental uncertainty problem, the dominant approach to fisheries management today is still this kind of “command and control” regime that tries to maintain the competitive system by, for example, determining shortened harvest seasons and caps on total allowable catch. However, even if the state succeeds in enforcing a sustainable limit of total catch, experience shows that maintaining competitive conditions in a common pool resource can be uneconomical and irrational. The reason is that without further technological-institutional support,

such a system leads to a rush for fish where each fisherman tries to maximize his catch, and a need to invest in the latest technology that leads to heavily overcapitalized fleets. This means an extreme duplication of material and a wasteful excess capacity so that the whole industry is operating at greater costs than benefits. In addition, the whole annual amount of total allowable catch is brought ashore in just a few days, and the race-like process leads to the waste of a high percentage of the caught fish.

In several instances this dilemma triggered the formation of cooperatives in fishing communities to reduce overexploitation and duplication. Paradoxically, the competition of cheap farmed salmon has also supported this trend towards sharing profits and efforts in co-ops to become more cost-efficient (Leal 2005: ch. 1; Wilen 2005: 50, 64–5). As a result of anti-trust law and opposition to supply curtailment, however, some governments impede self-organization to the detriment of ecologically beneficial arrangements (Adler 2005: 153). Anti-trust law is a legal institution that is used to prevent the creation of ecologically important associations. Ecology is not valued in the deliberation process (Adler 2005: 165). In places where anti-trust law is effective, co-ops have been allowed to prosper only where the state had already enacted strict catch limits that limited quantity regardless of the co-ops (Adler 2005: 162).

Limiting the spread of co-ops is unfortunate from a normative EE standpoint since more complex forms of co-management between the government and co-ops, and community-based management regimes (CBMR) often reflect an attempt to save the eco-system (Rose 2002: 250–1; Yandle 2005: 216–17). Another advantage of CBMR is that they successfully solve the problem of “command and control” regimes that caused the fishermen to only keep large fish and waste by-catch, ignoring important interspecies relations. “Co-management is a continuous problem-solving process, rather than a fixed state, involving extensive deliberation, negotiation and joint learning in problem-solving networks” (Plummer and Armitage 2007: 70). These are similar to traditional systems because they allow for resource-related variations and have turned resource conservation into an adaptive system (Rose 2002: 244), as, for instance, in New Zealand where the total allowable catch is defined in terms of percentage of resource (Rose 2002: 242) based on experience from historical practices and interactions with the resource base. Where CBMR has been enacted the role of the state has been to help facilitate solutions, such as limiting entry or fostering producer agreements (Townsend 2005: 142–5).

## **Conclusion**

This chapter underlines the renewed relevance of Kapp’s and Myrdal’s CCC by identifying common characteristics between CCC and the pre-analytical vision of EE in the area of fisheries. This propels the necessary integration of socio-economic and ecological knowledge based on a world view of real interrelations as well as an awareness of cumulative causation that can jeopardize societal reproduction and the satisfaction of fundamental human needs (Kapp 1961). It

also shows that CC theories of increasing returns must be embedded into the substantive CCC approach to elucidate whether or not the wealth-diminishing effects of social costs are so large that they actually offset gains in pecuniary increasing returns. In addition, Kapp's CCC-based analysis of salmon depletion and remedies for social costs are broadly compatible with EE. Therefore, the chapter supports an integrated ecological-institutional approach and at the same time advocates a greater role for the political economy of Kapp and Myrdal concerned with democracy and economic power.

## Notes

- 1 This transfer of fish protein away from protein-poor countries to already protein-rich countries for purposes of farming is described in Kapp (1976).
- 2 Galtung focuses on the exploitation of the periphery by the center to explain the depletion of fisheries and the ensuing regional poverty (Galtung 1975).
- 3 The consequences of resource depletion for social reproduction and unequal development are addressed by holistic economists, taking a socio-economic and ecological view of the development process (Galtung 1975), or focus on distribution conflicts (Guha and Martinez-Alier 1997).
- 4 Two important controversies center around Georgescu-Roegen's application of the entropy law to economics and his views on agrarian economics. The latter controversy is rooted in the conflict between Narodniki and Marxists (Guha and Martinez-Alier 1997: 25; Patnaik 1979).

## References

- Adaman, F., Devine, P. and Ozkaynak, B. (2003) "Reinstituting the Economic Process: (Re)embedding the Economy in Society and Nature," *International Review of Sociology*, 13, 2: 357–72.
- Adler, J. H. (2005) "Antitrust Barriers to Cooperative Fishery Management," in D. R. Leal (ed.) *Evolving Property Rights in Marine Fisheries*, Oxford: Rowman & Littlefield Publishers.
- Barnes, W. (2006) "Ecological and Institutional Economics: Building on Common Ground," paper presented at the Annual Meeting of the Association for Institutional Thought, Phoenix, April.
- Barthelemy, D. and Nieddu, M. (2007) "Non-trade Concerns in Agriculture and Environmental Economics: How J. R. Commons and Karl Polanyi can help us," paper presented at the Annual Meeting of the Association for Evolutionary Economics, Chicago, January.
- Berger, S. (2008a) "K. William Kapp's Theory of Social Costs and Environmental Policy: Towards Political Ecological Economics," *Journal of Ecological Economics*, 67: 244–52.
- (2008b) "Karl Polanyi's and K. William Kapp's Substantive Economics: Important Insights from the Polanyi–Kapp Correspondence," *Review of Social Economy*, 66, 3: 381–96.
- Berger, S. and Elsner, W. (2007) "European Contributions to Evolutionary Institutional Economics: The Cases of Cumulative Circular Causation (CCC) and Open System Approach (OSA) – Some Methodological and Policy Implications," *Journal of Economic Issues*, 41, 2: 529–37.

- Berger, S. and Forstater, M. (2007) "Towards Political Institutional Economics. Kapp's Social Costs, Lowe's Instrumental Analysis, and the European Institutional Approach to Environmental Policy," *Journal of Economic Issues*, 41, 2: 539–46.
- Clay, J. (2004) *World Agriculture and the Environment: A Commodity by Commodity Guide to Impacts and Practices*, Washington, DC: Island Press.
- Daughton, C. G. and Ternes, T. A. (1999) "Pharmaceuticals and Personal Care Products in the Environment: Agents of Subtle Change?" *Environmental Health Perspectives*, 107, 6: 907–38.
- De Groot, S. J. (2002) "A Review of the Past and Present Status of Anadromous Fish Species in the Netherlands: Is Restocking the Rhine Feasible?," *Hydrobiologia*, 478: 205–18.
- Dietz, T., Dolsak, N., Ostrom, E. and Stern, P. C. (2002) "The Drama of the Commons," in E. Ostrom, T. Dietz, N. Dolsak, P. C. Stern, S. Stonich and E. U. Weber (eds) *The Drama of the Commons*, Washington, DC: National Academy Press.
- Foran, J. A., Carpenter, D. O., Hamilton, M. C., Knuth, B. A. and Schwager, S. J. (2005) "Risk-based Consumption Advice for Farmed Atlantic and Wild Pacific Salmon Contaminated with Dioxins and Dioxin-like Compounds," *Environmental Health Perspectives*, 113, 5: 552–6.
- Foster, J. B. (2000) *Marx's Ecology: Materialism and Nature*, New York: Monthly Review Press.
- Galtung, J. (1975) "Development From Above and the Blue Revolution: The Indo-Norwegian Project in Kerala," in *International Peace Research Institute*, Oslo: PRIO Publication No. 2–12.
- Georgescu-Roegen, N. [1965] (1976) "Process in Farming vs. Process in Manufacturing: A Problem of Balanced Development," in *Energy and Economic Myths – Institutional and Analytical Economic Essays*, New York: Pergamon.
- (1970) "The Economics of Production," *American Economic Review*, 60, 2: 1–9.
- (1971) *The Entropy Law and The Economic Process*, Cambridge, MA: Harvard University Press.
- Guha, R. and Martinez-Alier, J. (1997) *Varieties of Environmentalism: Essays North and South*, London: Earthscan Publications.
- Helfield, J. M. and Naiman, R. J. (2001) "Effects of Salmon-derived Nitrogen on Riparian Forest Growth and Implications for Stream Productivity," *Ecology*, 82, 9: 2403–9.
- Hildebrand, G. V., Farley, S. D., Schwartz, C. C. and Robbins, C. T. (2004) "Importance of Salmon to Wildlife: Implications for Integrated Management," *Ursus*, 15, 1: 1–9.
- Kaldor, N. [1972] (1978) "The Irrelevance of Equilibrium Economics", in N. Kaldor (ed.) *Further Essays in Economic Theory*, New York: Holmes & Meier Publishers.
- (1978) "Introduction," in N. Kaldor (ed.) *Further Essays on Economic Theory*, New York: Holmes & Meier Publishers.
- [1974] (1978) "What is Wrong with Economic Theory," in N. Kaldor (ed.) *Further Essays on Economic Theory*, New York: Holmes & Meier Publishers.
- Kapp, K. W. (1950) *The Social Costs of Private Enterprise*, Cambridge, MA: Harvard University Press.
- (1961) *Toward a Science of Man in Society – A Positive Approach to the Integration of Social Knowledge*, The Hague: Martinus Nijhoff.
- [1965a] (1983) "Social Costs in Economic Development," in J. Ullman (ed.) *Social Costs, Economic Development and Environmental Disruption*, Lanham, MD: University Press of America.

- (1965b) "Social Economics and Social Welfare Minima," in T. K. N. Unnithan *et al.* (eds) *Towards a Sociology of Culture in India, Essays in Honor of Dr. D. P. Mukerji*, New Delhi: Prentice Hall of India.
- (1968) "In Defense of Institutional Economics," *Swedish Journal of Economics*, LXX, 1: 1–18.
- (1970a) "Environmental Disruption and Social Costs: A Challenge to Economics," *Kyklos*, 23, 4: 833–48.
- [1970b] (1983) "Environmental Disruption: General Issues and Methodological Problems," in J. Ullman (ed.) *Social Costs, Economic Development and Environmental Disruption*, Lanham, MD: University Press of America.
- (1972) "Social Costs, Neo-classical Economics, Environmental Planning: A Reply," *Social Science Information*, 11, 1: 17–28; reprinted in *The Social Costs of Business Enterprise* (1977), Nottingham: Spokesman.
- [1973] (1974) "Environmental Indicators as Indicators of Social Use Value," in K. W. Kapp (ed.) *Environmental Policies and Development Planning in Contemporary China and Other Essays*, Paris/The Hague: Mouton.
- (1976) "The Open-system Character of the Economy and its Implications," in K. Dopfer (ed.) *Economics in the Future*, London: Macmillan.
- (1950; 2nd edn 1963; 3rd edn 1977) *The Social Costs of Business Enterprise*, Nottingham: Spokesman University Paperback.
- Unpublished manuscript. *The Foundations of Institutional Economics*, Kapp Archive, Basel (CH).
- Knapp, G., Roheim, C. and Anderson, J. (2007) *The Great Salmon Run: Competition between Wild and Farmed Salmon*, Washington, DC: TRAFFIC North America, World Wildlife Fund. Online. Available at: [www.worldwildlife.org/trade/salmonreport.cfm](http://www.worldwildlife.org/trade/salmonreport.cfm) (accessed January 2007).
- Kocik, J. F. and Sheehan, T. F. (2006) "Atlantic Salmon," *Status of Fishery Resources off the Northeastern US*. NOAA Publication. Online. Available at: [www.nefsc.noaa.gov/sos/spsyn/af/salmon/index.html](http://www.nefsc.noaa.gov/sos/spsyn/af/salmon/index.html) (accessed January 2007).
- Leal, D. R. (2005) "Fencing the Fishery: A Primer on Rights-based Fishing," in D. R. Leal (ed.) *Evolving Property Rights in Marine Fisheries*, Oxford: Rowman & Littlefield Publishers.
- McGinnity, P., Prodöhl, P., Ferguson, A., Hynes, R., Maoiléidigh, N. O., Baker, N., Cotter, D., O'Hea, B., Cooke, D., Rogan, G., Taggart, J. and Cross, T. (2003) "Fitness Reduction and Potential Extinction of Wild Populations of Atlantic Salmon, *Salmo salar*, as a Result of Interactions with Escaped Farm Salmon," *Proceedings: Biological Sciences*, 270, 1532: 2443–50.
- Mercer, D. (2006) "Human Destruction of the Marine Commons: The Case of Fisheries Collapse in the North Atlantic," *GeoDate*, 19, 4: 5–7.
- Mullon, C., Fréon, P. and Cury, P. (2005) "The Dynamics of Collapse in World Fisheries," *Fish and Fisheries*, 6, 2: 111–20.
- Myers, R. A. and Worm, B. (2003) "Rapid Worldwide Depletion of Predatory Fish Communities," *Nature*, 423: 280–3.
- Myrdal, G. (1957) *Economic Theory and Under-developed Regions*, London: Gerald Duckworth.
- (1976) "The Meaning and Validity of Institutional Economics," in K. Dopfer (ed.) *Economics in the Future*, London: Macmillan.
- (1978) "Institutional Economics," *Journal of Economic Issues*, 12, 4: 771–83.



- Naylor, R. L., Goldburg, R. J., Mooney, H., Beveridge, M., Clay, J., Folke, C., Kautsky, N., Lubchenco, J., Primavera, J. and Williams, M. (1998) "Nature's Subsidies to Shrimp and Salmon Farming," *Science Magazine*, 282, 5390: 883–4.
- O'Hara, P. A. (2007) "Principles of Institutional-evolutionary Political Economy – Converging Themes from the Schools of Heterodoxy," *Journal of Economic Issues*, 41, 1: 1–42.
- Ostrom, E. (2005) *Understanding Institutional Diversity*, Princeton, NJ, and Oxford: Princeton University Press.
- (2006) "Bookreview – Institutions and the Environment, by Arild Vatn," *Land Economics*, 82, 2: 316–19.
- Özveren, E. (2007) "Where Disciplinary Boundaries Blur: The Environmental Dimension of Institutional Economics," in S. Ioannides and K. Nielsen (eds) *Economics and the Social Sciences: Boundaries, Interaction and Integration*, Cheltenham, and Northampton, MA: Edward Elgar.
- Paavola, J. and Adger, W. N. (2005) "Institutional Ecological Economics," *Journal of Ecological Issues*, 53: 353–68.
- Patnaik, U. (1979) "Neo-populism and Marxism: The Chayanovian View of the Agrarian Question and Its Fundamental Fallacy," *The Journal of Peasant Studies*, 6, 4: 375–420.
- Perroux, F. (1950) "The Domination Effect and Modern Economic Theory," *Social Research*, 17: 188–206.
- Plummer, R. and Armitage, D. (2007) "A Resilience-based Framework for Evaluating Adaptive Co-management: Linking Ecology, Economics and Society in a Complex World," *Journal of Ecological Economics*, 61: 62–74.
- Röpke, I. (2005) "Trends in the Development of Ecological Economics from the Late 1980s to the Early 2000s," *Journal of Ecological Economics*, 55: 262–90.
- Rose, C. M. (2002) "Common Property, Regulatory Property and Environmental Protection: Comparing Community-based Management to Tradable Environmental Allowances," in E. Ostrom, T. Dietz, N. Dolsak, P. C. Stern, S. Stonich and E. U. Weber (eds) *The Drama of the Commons*, Washington, DC: National Academy Press.
- Saunders, R., Hachey, M. A. and Fay, C. W. (2006) "Maine's Diadromous Fish Community: Past, Present, and Implications for Atlantic Salmon Recovery," *Fisheries*, 31, 11.
- Steppacher, R. (forthcoming) "Property, Mineral Resources and Sustainable Development," in O. Steiger (ed.) *Property Rights, Creditor's Money and the Foundation of the Economy*, Marburg: Metropolis.
- Stern, P. C., Dietz, T., Dolsak, N., Ostrom, E. and Stonich, S. (2002) "Knowledge and Questions after 15 years of Research," in E. Ostrom, T. Dietz, N. Dolsak, P. C. Stern, S. Stonich and E. U. Weber (eds) *The Drama of the Commons*, Washington, DC: National Academy Press.
- Swaney, J. A. (1985) "Economics, Ecology, and Entropy," *Journal of Economic Issues*, 19, 4: 853–65.
- (2006) "Policy for Social Costs: Kapp vs. Neo-classical Economics," in W. Elsner, P. Frigato and P. Ramazzotti (eds) *Social Costs and Public Action in Modern Capitalism – Essays Inspired by Karl William Kapp's Theory of Social Costs*, London and New York: Routledge.
- Swaney, J. A. and Evers, M. A. (1989) "The Social Cost Concepts of K. William Kapp and Karl Polanyi," *Journal of Economic Issues*, 23, 1: 7–33.
- Taylor, B. (1999) "Salmon and Steelhead Runs and Related Events of the Clackamas River Basin – A Historical Perspective," Portland General Electric Company. Online. Available at: [www.portlandgeneral.com/community\\_and\\_env/hydropower\\_and\\_fish/clackamas/history/clackamas\\_river\\_history\\_full.pdf](http://www.portlandgeneral.com/community_and_env/hydropower_and_fish/clackamas/history/clackamas_river_history_full.pdf) (accessed December 2008).

- Tietenberg, T. (2002) "The Tradable Permits Approach to Protecting the Commons: What Have We Learned?," in E. Ostrom, T. Dietz, N. Dolsak, P. C. Stern, S. Stonich and E. U. Weber (eds) *The Drama of the Commons*, Washington, DC: National Academy Press.
- Townsend, R. E. (2005) "Producer Organizations and Agreements in Fisheries: Integrating Regulation and Cosean Bargaining," in D. R. Leal (ed.) *Evolving Property Rights in Marine Fisheries*, Oxford: Rowman & Littlefield Publishers.
- Vatn, A. (2005) *Institutions and the Environment*, Cheltenham: Edward Elgar.
- Veblen, T. B. [1899] (1917) *The Theory of the Leisure Class – An Economic Study of Institutions*, New York: Macmillan.
- (1904) *The Theory of Business Enterprise*, New York: Charles Scribner's Sons.
- Wilens, J. E. (2005) "Property Rights and the Texture of Rents in Fisheries," in D. R. Leal (ed.) *Evolving Property Rights in Marine Fisheries*, Oxford: Rowman & Littlefield Publishers.
- Willson, M. F. and Halupka, K. C. (1995) "Anadromous Fish as Keystone Species in Vertebrate Communities," *Conservation Biology*, 9, 3: 489–97.
- Wilson, J. (2002) "Scientific Uncertainty, Complex Systems, and the Design of Common-pool Institutions," in E. Ostrom, T. Dietz, N. Dolsak, P. C. Stern, S. Stonich and E. U. Weber (eds) *The Drama of the Commons*, Washington, DC: National Academy Press.
- Worm, B., Barbier, E. B., Beaumont, N., Duffy, J. E., Folke, C., Halpern, B. S., Jackson, J. B. C., Lotze, H. K., Micheli, F., Palumbi, S. R., Sala, E., Selkoe, K. A., Stachowicz, J. J. and Watson, R. (2006) "Impacts of Biodiversity Loss on Ocean Ecosystem Services," *Science*, 314, 3: 787–90.
- Yandle, T. (2005) "Developing a Co-management Approach in New Zealand Fisheries," in D. R. Leal (ed.) *Evolving Property Rights in Marine Fisheries*, Oxford: Rowman & Littlefield Publishers.

## 10 Circular and cumulative causation in the classics

### Anticipations, family resemblances, and the influence on Post Keynesian economics

*Mathew Forstater and Michael J. Murray*

The origins of the theory of circular and cumulative causation are complex and diffuse, but are to be found mainly in the classical tradition of political economy.  
(Toner 1999)

#### Introduction

The theory of circular and cumulative causation is receiving renewed attention (see, e.g., Berger 2008; Forstater 2004). Early proponents of the principle were Veblen (1898) and Allyn Young (1928), with later elaborations by Myrdal (1957) and Kaldor (1967). Another early, and underappreciated, expositor of the concept was Adolph Lowe (Lowe [1935] 2003; Forstater 2004). In addition, it is not clear that the other writers directly influenced Lowe in this regard. Like Young and Kaldor, Lowe found inspiration for the notion in Adam Smith, but for Lowe it was rather the work of the classical economists and Marx generally that exhibited the basic vision of cumulative processes.

For Lowe, a key difference between the classical and neoclassical approaches regards “the entire possible *range* of deductive reasoning”:

Let us be quite clear about the disputed region. It concerns the entire natural, social, and technical environment of the economic system ... the changes in these elements through time ... [For the classical economists and Marx] the explanation of the order and changes of these data formed part of the theoretical work of economists.

(Lowe 1954: 109; emphasis added)

To put it into contemporary terms, for the classics and Marx many more components of the economic system – and even its natural and social environment – are treated as *endogenous*, that is, determined by forces internal to the system itself. Lowe even went so far as to argue that “upon this issue of endogeneity versus exogeneity, rather than conflicting theories of value, hinges the main difference between genuine classical theory and post-Millian economic reasoning, including all versions of neoclassical analysis” (Lowe 1954: 108).

Thus, three of the characteristics we will be looking for when mining the writings of the classics and Marx for anticipations of and family resemblances to the notion of circular and cumulative causation will be endogenous growth, endogenous technical change, and so-called “non-economic” factors as part of the analytical subject matter. Other factors – what we might call characteristics of circular and cumulative causation – are: increasing returns (especially in the manufacturing sector); external economies (in particular, pecuniary ones); positive feedbacks; backward and forward intersectoral linkages; structural change; learning-by-doing; and historical (rather than logical or notional) time. This list is not exhaustive, and not all of these factors need to be present in order for the links to be made with circular and cumulative causation.

The remainder of this chapter will proceed as follows. Next, an examination will be made of some examples of the theories of Smith, Ricardo and Marx with anticipations of and family resemblances to the principle of circular and cumulative causation. Following that, an investigation will be made of the ideas of two Post Keynesian economists, Luigi Pasinetti and Edward Nell, whose works have been influenced by the classics and Marx, and that exhibit some of these same characteristics of circular and cumulative causation.

## **Circular and cumulative causation in the classics and Marx**

### ***Adam Smith***

Of all the classical economists, Smith is the one most associated with contributing to the development of the notion of circular and cumulative causation. These ideas are primarily found in the first three chapters of Book I of *An Inquiry into the Nature and Causes of the Wealth of Nations* (Kaldor famously remarked that economics went wrong in the fourth chapter of Book I of *The Wealth of Nations*!). As is well known, it is in these initial chapters of Smith’s most famous work that he discusses how the division of labor increases productivity, the most important factor in determining a nation’s per capita output. But there is much, much more, both in these chapters and in subsequent ones, insights that comprise Smith’s dynamic view of capitalism’s growth and development.

After outlining his famous example of how the division of labor in a pin factory permits a fantastic increase in output per worker (i.e., increasing returns), Smith notes that “The nature of agriculture, indeed, does not admit of so many subdivisions of labour, nor of so complete a separation of one business from another, as manufactures”:

This impossibility of making so complete and entire a separation of all the different branches of labour employed in agriculture, is perhaps the reason why the improvement of the productive powers of labour in this art, does not always keep pace with their improvement in manufactures. The most opulent nations, indeed, generally excel all their neighbours in agriculture as

well as in manufactures; but they are commonly more distinguished by their superiority in the latter than in the former.

(Smith 1986: 163)

This emphasis on dynamic potential of the manufacturing sector, in particular, would later be taken up by Young, Kaldor, and others (see, e.g., Ricoy 1987).

Development of a nation's manufacturing sector as a whole leads to increasing productivity through external economies and increasing returns, in turn leading to a competitive advantage in global markets. A positive cycle of cumulative causation is engendered as competitive success increases aggregate demand, which leads to another round of growth and productivity increases. Such success is reflected on a nation's balance-of-payments position. On the other hand, sluggish growth in a nation's manufacturing sector means slow rates of productivity increases and weakness in global markets, reflected in the balance of payments. This latter scenario is typical of those nations assigned the role of primary-product producers in the international division of labor. Specialization in primary production means low income elasticities of demand and leaves the scope for technical change in agriculture dependent on importing capital. Industrialized countries benefit from the internal dynamics of the manufacturing sector with high income elasticities of demand for dynamic industries, harkening back to Lowe's emphasis on the machine tools sector (Lowe 1976; Argyrous 1996). The splitting up of the world into competitively successful manufacturing economies and sluggish primary economies was dubbed by Kaldor (1981) the "polarization process" and has important points of contact with the Prebisch-Singer thesis regarding the declining net barter terms of trade for developing economies and the structuralist interpretation of the world economic system.

Following the discussion of the three ways in which the division of labor increases productivity, Smith notes the great amount of other workers' labor embodied even in one cheap coat, and therefore the great amount of cooperation required by a social economy based on a system of general specialization. It is worthwhile to examine this passage at length:

Observe the accommodation of the most common artificer or day-labourer in a civilized and thriving country, and you will perceive that the number of people of whose industry a part, though but a small part, has been employed in procuring him this accommodation, exceeds all computation. The woollen coat, for example, which covers the day-labourer, as coarse and rough as it may appear, is the produce of the joint labour of a great multitude of workmen. The shepherd, the sorter of the wool, the wool-comber or carder, the dyer, the scribbler, the spinner, the weaver, the fuller, the dresser, with many others, must all join their different arts in order to complete even this homely production. How many merchants and carriers, besides, must have been employed in transporting the materials from some of those workmen to others who often live in a very distant part of the country! How much com-

merce and navigation in particular, how many ship-builders, sailors, sail-makers, rope-makers, must have been employed in order to bring together the different drugs made use of by the dyer, which often come from the remotest corners of the world! What a variety of labour too is necessary in order to produce the tools of the meanest of those workmen! To say nothing of such complicated machines as the ship of the sailor, the mill of the fuller, or even the loom of the weaver, let us consider only what a variety of labour is requisite in order to form that very simple machine, the shears with which the shepherd clips the wool. The miner, the builder of the furnace for smelting the ore, the feller of the timber, the burner of the charcoal to be made use of in the smelting-house, the brick-maker, the brick-layer, the workmen who attend the furnace, the mill-wright, the forger, the smith, must all of them join their different arts in order to produce them. Were we to examine, in the same manner, all the different parts of his dress and household furniture, the coarse linen shirt which he wears next his skin, the shoes which cover his feet, the bed which he lies on, and all the different parts which compose it, the kitchen-grate at which he prepares his victuals, the coals which he makes use of for that purpose, dug from the bowels of the earth, and brought to him perhaps by a long sea and a long land carriage, all the other utensils of his kitchen, all the furniture of his table, the knives and forks, the earthen or pewter plates upon which he serves up and divides his victuals, the different hands employed in preparing his bread and his beer, the glass window which lets in the heat and the light, and keeps out the wind and the rain, with all the knowledge and art requisite for preparing that beautiful and happy invention, without which these northern parts of the world could scarce have afforded a very comfortable habitation, together with the tools of all the different workmen employed in producing those different conveniences; if we examine, I say, all these things, and consider what a variety of labour is employed about each of them, we shall be sensible that without the assistance and co-operation of many thousands, the very meanest person in a civilized country could not be provided, even according to what we very falsely imagine, the easy and simple manner in which he is commonly accommodated.

(Smith 1986: 167–8)

This wonderful description demonstrates the crucial *interdependencies* emphasized by later theorists of circular and cumulative causation, so related to Young's insistence that the most powerful cumulative processes are "macro-phenomena." But what has not been recognized, perhaps, about this important passage is that it is also a terrific description of backward intersectoral linkages (Hirschman 1958)!

Interestingly, Darity (1992), following the Williams–Rodney thesis, has employed the notions of backward and forward intersectoral linkages introduced by Hirschman in another version of the polarization process, "the rise of the west and the lag of the rest." Enslaved Africans laboring on the Native American lands produced otherwise impossibly cheap raw materials and food for Europe. This lessened

agricultural requirements in Europe and supplied the raw materials for incipient European manufactures. The manufactured goods, made with raw materials produced on plantations by a labor force fed on plantation agricultural produce, were then taken to Africa, interrupting traditional crafts, and further financing industry back in Europe and plantations in the Americas. This is the famous Triangular Trade.

What kinds of goods and industries were involved in this process? – sugar, spices, cocoa, coffee, cotton, tea. Sugar, however, means not only sugar cane, but gives birth to refineries producing refined sugar and molasses and rum. This creates demand for all the tools and implements and machines used in growing, harvesting, processing, and transporting these goods. Furthermore, the production of these tools and implements and machines itself requires other tools and implements and machines, as well, of course, as labor. The labor employed in these new industries also stimulates consumption demand, feeding into a whole host of consumption goods industries that require tools and implements and machines, and more labor, for their production, and on and on. In addition, there is the role of the shipbuilding industry and gun manufacturing, in the same manner giving rise to industries producing and servicing their component inputs. This frenzy of activity led to what Smith and later Marx called the primary or primitive accumulation necessary for capitalism's "take-off." At the same time, it also caused a depopulation of Africa's youngest and most productive inhabitants, preventing the development of a widespread division of labor and other "salutary effects Adam Smith foresaw for a region anticipating growth in markets" (Darity 1992: 165). In addition, it was not only sugar and cotton in Africa; the fur trade in Canada and Siberia and spices and tea in India must be considered to fathom the scale and impact (the European discovery of the sea-route to Southern Asia around the Cape of Good Hope occurred in the same year as Columbus first arrived in the Americas).

With regards to this primary accumulation, Smith argued that a precondition for the division of labor is that some previous accumulation of capital must have taken place. Further, Smith underscores the fact that not only is the accumulation of capital a necessary precondition for the division of labor, it also leads to the improvement of the division of labor:

As the accumulation of stock must, in the nature of things, be previous to the division of labour, so labour can be more and more subdivided in proportion only as stock is previously more and more accumulated. ... As the accumulation of stock is previously necessary for carrying on this great improvement in the productive powers of labour, so that accumulation naturally leads to this improvement.

(Smith 1986: 228)

Young noted that this might sound like a circular argument, but that it is much more than that; it is the mutually reinforcing relation between economic growth and technical change. This fundamental, powerful dynamic is at the heart of Smith's theories of endogenous growth and endogenous technical change.



As Lowe (1965), Heilbroner (1973), and others have argued, Smith's growth model may be envisioned as a spiral, a circular process that continues on an expanding scale until it pushes up against some ultimate constraints. Once the historical institutions are in place and primitive accumulation has occurred, the analysis can be entered at any stage. We begin with capitalist investment, rooted in the desire of "bettering one's condition," itself the result of the need to be the object of "sympathy" (i.e., empathy), as Smith argued in *The Theory of Moral Sentiments*, and which is ultimately a "deception," since it often does not pay off (Smith 1986: 120–1, 135).

Because accumulation involves capital and labor, and accumulation is the precondition for a greater division of labor, demand for labor increases. Increased labor demand puts an upward pressure on wages and, if this were the end of the story, accumulation might cease since higher wages mean higher costs for capitalists, which cut into profits. This is not what occurs, however, because here Smith introduces what constitutes another important endogenous dynamic of the classical approach: "the demand for men, like that for any other commodity, necessarily regulates the production of men" (Smith 1986: 204–5). Higher wages increase population, increasing labor supply, and causing wages to fall. With higher employment and wages at least as high as they were initially (and perhaps a little higher), the wage bill will be greater. With a workforce with greater purchasing power, consumption demand will be higher; in other words, the market will be bigger. Bigger markets and more demand mean more sales and profits, meaning more accumulation and a more refined division of labor (greater subdivision of tasks, introduction of machinery, higher productivity). Thus growth continues to spiral outward.

Here we have introduced another of the internal dynamics, one at least as important as those regarding labor supply and productivity. The famous title of Chapter III of Book I of the *Wealth of Nations*, "That the division of labour is limited by the extent of the market," was called by Young "one of the most illuminating and fruitful generalisations which can be found anywhere in the whole literature of economics" (Young 1928: 529):

As it is the power of exchanging that gives occasion to the division of labour, so the extent of this division must always be limited by the extent of that power, or, in other words, by the extent of the market. When the market is very small, no person can have any encouragement to dedicate himself entirely to one employment, for want of the power to exchange all that surplus part of the produce of his own labour, which is over and above his own consumption, for such parts of the produce of other men's labour as he has occasion for. There are some sorts of industry, even of the lowest kind, which can be carried on nowhere but in a great town. A porter, for example, can find employment and subsistence in no other place. A village is by much too narrow a sphere for him; even an ordinary market town is scarce large enough to afford him constant occupation.

(Smith 1986: 171)

The flip-side of this, of course, is that bigger markets permit greater specialization. Interestingly, Smith notes much later on, in Book IV, Chapter I, that the main economic benefit to Europe of the discovery of the Americas was not the gold and silver obtained there, but the dynamic set in place by the opening up of new and bigger markets:

By opening a new and inexhaustible market to all the commodities of Europe, it gave occasion to new divisions of labour and improvements of art, which, in the narrow circle of ancient commerce, could never have taken place for want of a market to take off the greater part of that produce. The productive powers of labour were improved, and its produce increased in all the different countries of Europe, and together with it the real revenue and wealth of the inhabitants.

(Smith 1937: 415–16)

The sentence that follows this passage is most interesting, given the tendency of some authors to claim that Smith recognized process innovations, but not product innovations (see, e.g., Heilbroner 1973): “The commodities of Europe were almost all new to America, and many of those of America were new to Europe” (Smith 1937: 416). Of course, improved product quality and design and not only lower costs are an outgrowth of advancing productivity.

Given that Kaldor (1972: 1240–1) has stated not only that economics went wrong beginning in Chapter IV of Book I of *The Wealth of Nations*, but specifically refers to Smith’s notion of natural price (introduced in Book I, Chapter VII), it seems relevant to briefly note how, perhaps *contra* Kaldor, this and related ideas might also be seen as part of the classical dynamics. In addition, the distinction between natural and market prices (and their relation) is made not only by Smith, but by Ricardo and Marx as well. The market price of a commodity can deviate from its natural price when there are imbalances between quantity supplied and effectual demand. These imbalances set a process in motion such that market price will tend to the natural price. The classical notion of competition states that capital seeking the highest rate of return will tend to establish a uniform rate of profits between industries. This dynamic process has in modern versions of classical analysis been termed cross-dual dynamics, which Flaschel and Semmler have expressed as:

(i) the output of a commodity is expanded or reduced (through entry or exit of firms) whenever the excess of price over cost (including normal profits) is positive or negative (“Law of excess profitability”); and, (ii) the price of a commodity is raised or lowered whenever there is an excess demand or supply on the market (“Law of excess demand”).

(Flaschel and Semmler 1988: 3)

Recent discussions and formalizations demonstrate that far from some simple, static framework, such an adjustment process represents a dynamic and nonlinear

relationship between supply and demand imbalances, relative price changes, profit rate differentials, and changes in sectoral growth rates, and “must surely be regarded as one of the richest legacies of [Smith’s] thought, and of the Classical tradition” (Walsh 1992: 14).

### **David Ricardo**

Ricardo is not often mentioned in the context of the principle of circular and cumulative causation, but not only do you find in his work such cross-dual dynamics as we have just discussed, but endogenous growth and structural change as well. While Smith, as we have seen, emphasized increasing returns in manufacturing, Ricardo focused on diminishing returns in agriculture. In both his *Essay on Profits* and his *On the Principles of Political Economy and Taxation*, Ricardo considers what will happen in an advancing capitalist economy depending on natural resources such as land in the course of its development. He begins with the following basic vision: “This accumulation would lead to an increased demand for labour, to higher wages, to an increased population, to a further demand for raw produce, and to an increased cultivation” (Ricardo 1951: 790). Here, we see that Ricardo accepts the classical theory of population, already introduced in the discussion of Smith above. Ricardo, however, then takes this in another direction, as he raises the issue of how the larger population leads to an increase in the demand for food and other primary commodities. This prompts Ricardo to introduce a distinctly classical notion of diminishing returns.

Ricardo’s famous Chapter II of his *Principles*, “On Rent,” contains his theory of differential rent, first introduced by him in his *Essay on Profits*, and also by some of his contemporaries, such as Malthus and Torrens.

On the first settling of a country, in which there is an abundance of rich and fertile land, a very small proportion of which is required to be cultivated for the support of the actual population, or indeed can be cultivated with the capital which the population can command, there will be no rent; for no one would pay for the use of land, when there was an abundant quantity not yet appropriated, and, therefore, at the disposal of whosoever might choose to cultivate it.

(Ricardo 1951: 69)

However, land is not available in infinite supply nor is all land of equal fertility:

It is only, then, because land is not unlimited in quantity and uniform in quality, and because in the progress of population, land of an inferior quality, or less advantageously situated, is called into cultivation, that rent is ever paid for the use of it. When in the progress of society, land of the second degree of fertility is taken into cultivation, rent immediately commences on that of the first quality, and the amount of that rent will depend on the difference in the quality of these two portions of land.

(Ricardo 1951: 70)

Here it is worthwhile to note how Ricardo's distinctly classical notion of diminishing returns differs from the neoclassical notion. Whereas in the classics, as we have seen, there is an asymmetric analysis of diminishing returns in agriculture and mining only, neoclassical models portray all factors of production symmetrically, experiencing increasing returns as production begins and diminishing returns as production continues (as in the U-shaped average cost curve). Neoclassical economics also abandoned the qualitative aspects of returns, limiting the causes to quantitative differences only. In neoclassical economics all factors of production are homogeneous, and so there are only diminishing returns of the intensive variety, in the classics diminishing returns may be extensive or intensive. Ricardo (1951: 70) even remarked on the possibilities of diminishing returns due to "peculiar advantages of situation," such as proximity to the central market outlets, resulting in differential transportation costs (an idea first introduced by William Petty, who Marx considered the originator of classical political economy). These asymmetrical and qualitative aspects of the classical approach bear family resemblances to the notion of circular and cumulative causation rather than the purely quantitative equilibrium approach of neoclassical economics.

Continuing with Ricardo's scenario, as accumulation takes place, population grows and the demand for food increases, cultivation will have to be extended on to less fertile land, where the productivity of labor is therefore less. If it is assumed, as a first approximation, that wages are still the same, profits must be less on the less fertile land. But profits cannot be less – competition enforces a tendency toward uniformity in the rate of profit. Therefore, the surplus on the least productive land determines the rate of profit throughout agriculture. Landlords with more productive land charge rent for its utilization. So the "super-profit" on the more productive land becomes rent. Meanwhile, though, as cultivation continues to be extended, profits are shrinking.

Following Ricardo's approach in his *Essay on Profits*, we now relax the assumption that wages remain constant. In his Chapter V of the *Principles*, "On Wages," Ricardo remarks that "with every improvement of society, with every increase in its capital, the market wages of labour will rise" (1951: 96). Just as in the case of commodities, the market wage will tend toward the natural wage, but even though the natural wage is determined by subsistence, in the classical theory wages are historically and socially – not biologically – determined by subsistence. Ricardo writes:

It is not to be understood that the natural price of labour, estimated even in food and necessaries, is absolutely fixed and constant. It varies at different times in the same country, and very materially differs in different countries. It essentially depends on the habits and customs of the people. An English labourer would consider his wages under their natural rate, and too scanty to support a family, if they enabled him to purchase no other food than potatoes, and to live in no better habitation than a mud cabin; yet these moderate demands of nature are often deemed sufficient in countries where "man's life is cheap," and his wants easily satisfied. Many of the conveniences now

enjoyed in an English cottage, would have been thought luxuries at an earlier period of our history.

(Ricardo 1951: 96–7)

Since the “market rate [of wages] may, in an improving society, for an indefinite period, be constantly above” their natural rate, rather than the market rate tending back down to the natural rate, the natural rate may in these circumstances end up rising to the market rate (1951: 94–5). In any case, the rising unit cost or cost per bushel of corn due to deteriorating production conditions (less fertile land) leads to a rise in the price of corn and thus increases the value of the means of subsistence. Profits get squeezed from below by wages as well as from above by rent. When profits fall, accumulation is choked off, investment is not taking place, and stagnation sets in.

There are two factors emphasized by Ricardo that can fend off the stationary or declining state: technological progress or foreign trade. While growth is endogenous in Ricardo, technological change is not. Yet, Ricardo in the third edition of the *Principles*, did add a chapter, “On Machinery,” in which he reversed an earlier position and put forward an analysis that recognized technological unemployment. In other words, Ricardo conceded the possibility that labor displaced by technical progress might not be employed elsewhere, which was the prevalent view up until that time. Ricardo’s analysis in Chapter XXXI has been called an “early and rude type of traverse analysis” (Kurz 1984). The traverse describes an economic system shunted off a steady-state growth path by *structural change*.

While macro-economics regards changes in the levels and rates of growth of aggregate economic activity, structural change deals with variations in the *composition* of economic activity and the effect such variation has on the operation of the economic system. Structural analysis thus often focuses on such factors as organizational and institutional evolution, changes in labor supply and the supply of natural resources, capital- and labor-displacing technological progress, and the composition of final demand. The importance of these factors for sector proportionality and balance means that structural analysis is often conducted at a slightly lower level of aggregation than most macro-economic models, highlighting (for example) inter-industry relations. Structural change may also be seen as changes in the *data* of economic theory and is therefore important for understanding the long-term development of the economic system. A nation’s structural position in the global economy often has important implications for such development, as well as for human, social, and environmental well-being generally. Structural change was an important part of the work of the classical political economists and Marx, and is at the heart of the principle of circular and cumulative causation.

### **Karl Marx**

Ricardo’s conclusion that workers displaced by technical progress might not be re-employed elsewhere in the economy was a point picked up and developed by Marx and incorporated into his analysis of growth and cycles. In Marx, however,

labor-displacing technological change is endogenous. Marx also outlined the requirements for proportionality and balance in his schemes of reproduction, with output in the capital goods sector replacing the fixed capital in both the capital and consumption goods sectors, and output in the consumption goods sector providing the means for subsistence for workers in both sectors. Expanded reproduction thus requires not only aggregate but also inter-sectoral proportionality and balance.

In Chapter 25 of Volume 1 of *Capital*, “The General Law of Capitalist Accumulation,” Marx lays out the dynamic relationship between unemployment, wages, profitability, and technical change. This is Marx’s theory of the trade cycle, but it also concerns the longer term rhythms of accumulation. Unlike many monetary theories of the business cycle, in Marx’s structural theory the same forces that determine the long-run trajectory of the system also explain the business cycle. Slumps may be exacerbated by monetary disturbances, but at bottom are regulated by the structural and technological features of capitalist accumulation.

Marx’s analysis in Chapter 25 commences in a manner reminiscent of both Smith and Ricardo, as outlined above: accumulation increases the demand for labor, causing wages to rise:

Growth of capital implies growth of its variable constituent, in other words, the part invested in labour-power. A part of the surplus-value which has been transformed into additional capital must always be re-transformed into variable capital, or additional labour-fund. If we assume that, while all other circumstances remain the same, the composition of capital also remains constant (*i.e.* that a definite mass of the means of production continues to need the same mass of labour-power to set it in motion), then the demand for labour, and the fund for the subsistence of the workers, both clearly increase in the same proportion as the capital, and with the same rapidity. Since the capital produces a surplus-value every year, of which one part is added every year to the original capital; since this increment itself grows every year along with the augmentation of the capital already functioning; and since, lastly, under conditions especially liable to stimulate the drive for self-enrichment, such as the opening of new markets, or of new spheres for the outlay of capital resulting from newly developed social requirements, the scale of accumulation may be suddenly extended merely by a change in the proportion in which the surplus-value or surplus product is divided into capital and revenue – for all these reasons the requirements of accumulating capital may exceed the growth in labour-power or in the number of workers; the demand for labourers may outstrip the supply, and thus wages may rise. This must indeed ultimately be the case if the conditions above continue to prevail. For since in each year more workers are employed than in the preceding year, sooner or later a point must be reached at which the requirements of accumulation begin to outstrip the customary supply of labour, and a rise of wages therefore takes place.

(Marx 1990: 763)

Here, however, Marx departs from the earlier discussions in the classics in some important ways. Recall that in both Smith and Ricardo, the increasing demands for labor and rising wages result in increased labor supply via the classical or Malthusian law of population. Marx harshly criticizes and roundly rejects any notion of “natural” or universal laws of population growth. Instead, he views “laws” of population as only tendencies of a particular mode of production. In capitalism, this takes the form of what Marx terms the *relative surplus population*, or *reserve army of labor*:

This is a law of population peculiar to the capitalist mode of production; and in fact every particular historical mode of production has its own special laws of population, which are historically valid within that particular sphere. An abstract law of population exists only for plants and animals, and even then only in the absence of any historical intervention by man.

(Marx 1990: 783–4)

For both Marx and Keynes, unemployment is a normal feature of capitalism, but whereas for Keynes unemployment is irrational, for Marx it is *functional*. Unemployment is not an irrational by-product; it serves several important purposes in the system. First, unemployment provides a pool of labor standing ready to work when the demand for labor rises in response to expansion. Second, unemployment helps to hold down wages by decreasing the bargaining power of labor. Third, unemployment disciplines workers, who will be fearful of dismissal when there are many potential replacements needing employment.

Another distinctive feature of Marx’s theory of endogenous growth is that it is driven by endogenous labor-displacing technical change. In Smith, there is endogenous technical change, but it is not labor-displacing; in Ricardo, there is labor-displacing technical change, but it is not endogenous. In Marx, competitive pressures require capitalists to seek to raise productivity and cut costs. Both productivity and costs are partially determined by technology.

We now have the pieces required to put together Marx’s theory. As capitalist expansion takes place, firms increase their demand for labor, wages get bid up, which are costs for business that cut into profits. Firms respond by instituting cost-cutting technical change. This technical change is labor-displacing, so unemployment rises and the reserve army expands, dampening wages, and at the same time increasing productivity and profits. As profits expand, growth expands, the demand for labor rises, and this puts upward pressure on wages, cutting into profits, and growth slackens, leading to another round of cost-cutting, labor-displacing technical change. This is the expansion and contraction of the aggregate reserve army.

However, over time, the capital–labor ratio is rising; in other words, firms don’t go back to the old, more labor-intensive methods of production when they increase their demand for labor, they are increasing their demand for labor with the new technical organization. So less labor is being used per machine as a secular trend. So, technical change alters the capital structure; more labor is



being used, but with the new production methods. While individual capitalists benefit from lower costs, the higher capital intensity of production results in a falling average rate of profit for the economy as a whole.

Not only does Marx have a theory of endogenous growth and a theory of endogenous technical change, he has a theory of endogenous structural change. In addition, there are a number of other features of Marx's system that have family resemblances with the principle of circular and cumulative causation.

Modern developers of the principle of circular and cumulative causation have often looked to Adam Smith for anticipations of the notion, but the theories of Ricardo and Marx also contain the basic vision of cumulative processes. Some contemporary Post Keynesian economists who have been strongly influenced by the classics and Marx have put forward frameworks that include, explicitly or implicitly, circular and cumulative causation. In the final section we will briefly examine two examples.

### **Transformational growth and structural economic dynamics**

Two Post Keynesian economists who take much of their inspiration from classical political economy (and Marx), and whose works contain important points of contact with the principle of circular and cumulative causation, are Edward J. Nell and Luigi L. Pasinetti. Both authors regard capitalism as a dynamic economic system exhibiting endogenous structural and technical change, and operating in historical time. This section briefly examines their ideas.

Nell's theory of transformational growth bears significant family resemblances to the notion of circular and cumulative causation. Contemporary capitalism exhibits ongoing structural and technological change, and so is not well depicted by instantaneous adjustments leading to equilibrium solutions.

Equilibrium theory is largely useless for practical purposes. It is unable to explain the ordinary behavior of business and the working of markets, let alone speculate on the economic forces that helped to create the modern world.

(Nell 1998: 3)

Transformational growth is the "interlocking emergence of new products and new processes, creating new markets, and new industries" (Nell 1992: 106). As in the classics and Marx, competition drives these changes.

Competitive market pressures especially breed innovation. Innovation leads to the expansion of existing markets, and to the creation of new markets, which in turn give rise to further competitive pressures, leading to a bigger push to innovate, furthering the expansion and the creation of markets, and so on. Transformational growth rejects the notion of steady-state growth, instead viewing growth as disproportional, qualitative, and disruptive. These kinds of transformative processes result in, *and are further caused by*, changes in socio-economic relationships, such as the "distribution of income, and the urban-rural relationship, together with the nature of work, of household life, and so on" (Nell

1988: 162). The process of transformational growth is an interdependent and cumulative process.

Nell employs transformational growth to examine the transition from a craft-based economy of family firms and family farms into a modern industrial mass production economy, the key features of which are not only production on a large scale, but production that is by its nature in continuous endogenous flux, and where economic processes are wholly interdependent. In a dynamic framework, economic data such as socioeconomic relationships, or parameters such as technical coefficients of production, are bound to change through time. As Lowe, a colleague and mentor of Nell at the New School for Social Research, has put it, once analysis moves beyond the short period into a dynamic framework, it is improper to treat economic data as given:

It is a fundamental theorem of realistic theory that under the particular social conditions of the industrial system, data and process are involved in a regular and continuous interaction which makes any concrete constellation, and therefore the system as a whole, essentially unstable and liable to transformation. For this reason in any long period analysis concerning the industrial system, on principle the data are to be handled as dependent variables.

(Lowe [1935] 2003: 146–7)

Pasinetti has also been influenced by Lowe's ideas, and was a student of Leontief, who was in residence at Kiel University in the 1920s while he wrote his doctoral dissertation on input–output analysis, and where Lowe was research director.

It should be noted that there are magnitudes, such as technology, or the attitudes and preference of consumers, that an economic analysis may consider as exogenously given; and yet they are extremely, *if not crucially*, variable over time.

(Pasinetti 1993: 11)

Structural analysis is indispensable for understanding the process of transformational growth. A structural model of production and distribution describes completely the: (1) industrial/sectoral relationships; (2) socio-technical relationships as expressed by final demand; and (3) conditions required for the system as a whole to reproduce itself. For the practical purposes of planning and policy, conventional input–output analysis may be the most appropriate to describe the interdependence of industries in a mass production economy. Models that are not as disaggregated as input–output models may be more useful for theoretical analysis, however. Horizontal models (such as Lowe's three-sector framework) emphasize sectoral interdependence, inter-industry linkages, and their associated implications (Lowe 1976). The horizontal models are advantageous, as they are parsimonious, and they highlight time lags, bottlenecks, and other rigidities which make up the physical and technical nature of the system. Growth of industries within a given sector create backward and forward linkages, as capital

goods are highly specific and are not shiftable between different lines of production. Industrial growth in non-primary sectors requires the formation of real capital goods, creating backward linkages, to be used as inputs into the expanding sector. Growth in primary sectors creates forward linkages, as increases in output are used both as inputs into the production of other goods, and a portion of the output is purchased by consumers, contributing to final demand.

The interdependencies of a mass production economy have circular and cumulative effects. The growth in sector  $i$  causes an increase in the demand for the output of sector  $j$ , which in turn increases income and employment in sector  $j$ , causing an increase in final demand, which creates positive feedbacks to sector  $i$ , generating further demand for output in sector  $i$ . Thus, any change in the final demand for goods and services in any given industry is bound to have a cumulative effect, since these changes happen in the same direction, increasing the magnitude of the result, such as in the multiplier processes of Keynes, central for both Nell and Pasinetti.

Inter-industry models such as input–output models or horizontally integrated models, with a given set of technical coefficients, are simply snapshots of the economic process at a given point in time. Horizontal models are useful for describing the physical, technical nature of the production system at a given point in time, and have been fruitfully employed to investigate the traverse from one growth path to another.

Pasinetti's structural economic dynamics replaces inter-industry relationships to a framework in terms of vertically integrated sectors. Pasinetti's framework recognizes that the economy is a multi-sectoral industrial system with ongoing technical change, and ongoing changes in the level and composition of final demand. Pasinetti's model, unlike the horizontally integrated models of Leontief and Lowe, is not linked by inter-industry coefficients, but rather each of the sectors is linked by the overall impacts of effective demand (see Pasinetti 1981, 1993, 2007). The nature of employment in Pasinetti's model is due to the division of labor among the sectors, which enables each individual laborer to contribute to only a small portion of the production process, but contributing to the demand of all the goods and services produced in the economy. Pasinetti's model, while not as parsimonious as the Lowe model, nor as descriptive as the Leontief input–output model, may be preferable for the investigation of certain problems, as the model has elements of effective demand (which affects the  $c_i$  coefficient) and technical progress (affecting the  $l_i$  coefficient).

The Pasinetti model describes the interdependencies of production and distribution, and the structural and technological requirements for the conditions for full employment.

#### PASINETTI'S MODEL

- (1) Movement of labor coefficients:  $l_i(t) = l_i(0)e^{-\rho t}$ — $i = 1, 2, \dots, m$
- (2) Movement of per capita consumption:  $c_i(t) = c_i(0)e^{\rho t}$
- (3) Population growth:  $N(t) = N(0)e^{gt}$

- (4) Structural dynamics:  $\rho_i \neq \rho_j$ ;  $r_i \neq r_j$ ;  $r_i \neq \rho_j$ —where:  $\rho_i$ ;  $r_i$ ;  $g \geq 0$   
 $N(t)$  = total population at time (t)  
 $l_i(t)$  = labor coefficients (labor required per unit of output at time t)  
 $\rho_i$  = rate of growth of the labor productivity in sector I  
 $c_i(t)$  = per capita consumption  
 $r_i$  = rate of change of the per capita consumption of commodity I  
 $g$  = rate of growth of the population  
 $\mu$  = proportion of active to total population  
 $v$  = proportion of working time to total time

From equations (1) and (2) above it can be seen that the condition for full employment is:

$$\frac{1}{\mu(t)v(t)} \times \sum_{i=1}^m c_i(0)l_i(0)e^{(\tau_i - \rho_i)t} - 1 = 0 \quad (5)$$

which states that the proportion of labor employed in each sector must add to one, and is dependent upon demand for output within each sector, as well as upon structural and technological change. The full employment condition is rarely if ever fulfilled and is certainly not a self-equilibrating process. The reason is due to the nature of the parameters  $r_i$ , which is essentially the effective demand condition, and  $\rho_i$ , labor productivity, which serves as a proxy for technical change. The overall effect is on sectoral employment,  $\xi_i$  ( $\xi_i = g + r_i - \rho_i$ ) which can be positive (absorbing labor into sector i) or negative (driving labor from sector i). If, as is most often the case, labor productivity is increasing in sector i, causing  $\rho_i$  to be positive, per capita consumption in sector j, k ... m must be

Pasinetti's Pure Labor Model: physical quantity system

$$\begin{bmatrix} 1 & 0 & \dots & 0 & \dots & 0 & -c_1(t) \\ 0 & 1 & \dots & 0 & \dots & 0 & -c_2(t) \\ \vdots & \vdots & \ddots & \vdots & & \vdots & \vdots \\ 0 & 0 & \dots & 1 & \dots & 0 & -c_i(t) \\ \vdots & \vdots & & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \dots & 0 & \dots & 1 & -c_m(t) \\ -l_1(t) & -l_2(t) & \dots & -l_i(t) & \dots & -l_m(t) & \mu(t)v(t) \end{bmatrix} \times \begin{bmatrix} Q_1(t) \\ Q_2(t) \\ \vdots \\ Q_i(t) \\ \vdots \\ Q_m(t) \\ N(t) \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ \vdots \\ 0 \\ \vdots \\ 0 \\ 0 \end{bmatrix}$$

Pasinetti's Pure Labor Model: price system

$$\begin{bmatrix} 1 & 0 & \dots & 0 & \dots & 0 & -l_1(t) \\ 0 & 1 & \dots & 0 & \dots & 0 & -l_2(t) \\ \vdots & \vdots & \ddots & \vdots & & \vdots & \vdots \\ 0 & 0 & \dots & 1 & \dots & 0 & -l_i(t) \\ \vdots & \vdots & & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \dots & 0 & \dots & 1 & -l_m(t) \\ -c_1(t) & -c_2(t) & \dots & -c_i(t) & \dots & -c_m(t) & \mu(t)v(t) \end{bmatrix} \times \begin{bmatrix} p_1(t) \\ p_2(t) \\ \vdots \\ p_i(t) \\ \vdots \\ p_m(t) \\ w(t) \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ \vdots \\ 0 \\ \vdots \\ 0 \\ 0 \end{bmatrix}$$

Figure 10.1 Pasinetti's pure labor model.

growing in order for labor to be reabsorbed and return the system to full employment. Pasinetti's result is that the economy, owing to its interdependence, is inherently unstable. Non-proportional sectoral growth has an effect on the entire economic system (Pasinetti 2007). Full employment is rarely if ever achieved, let alone maintained, due to the perpetual movement of per capita consumption and technical progress. The two are moving in opposite directions, and so will cancel each other out to some extent, "but never completely, in any systematic way" (2007: 286). Pasinetti shows quite clearly the complexity of the economic system. Industrial capitalism suffers from two types of unemployment: technological or Marxian unemployment and Keynesian unemployment, as described by Pasinetti's model. This leads Pasinetti to the following conclusion:

Keynes intuition is proved right and the implications are far reaching.... [The condition for full employment] is not a once and for all condition, except in the extreme trivial case of a perfectly stationary economic system.  
(Pasinetti 2007: 286)

In a dynamic system, every component of equation (5) is moving, corresponding changes in demand by consumers of given output in given sectors, which further has an effect on all other sectors of the economy. Structural models demonstrate that modern mass production economies are never at rest. They are in continuous flux, with the expansion of industries within sectors, the introduction of new product and process innovations, leading to the addition of new industries and the demise of obsolete, non-competitive industries. The role of the market in a mass production economy is to foster innovation. Markets reward winners through an increase in demand, and punish losers whose innovations do not succeed. Production is an interdependent process, and is by its nature unbalanced. Structural models are useful to illuminate the connection of industry to industry and sector to sector, and the connection of society to production, and production to society. Transformational growth, depicting structural and technological change, is an inherently dynamic, social, interdependent process, and dovetails nicely with the principle of circular and cumulative causation.

## References

- Argyrous, G. (1996) "Cumulative Causation and Industrial Evolution: Kaldor's Four Stages of Industrialization as an Evolutionary Model," *Journal of Economic Issues*, 30, 1: 97–119.
- Berger, S. (2008) "Circular Cumulative Causation (CCC) à la Myrdal and Kapp – Political Institutionalism for Minimizing Social Costs," *Journal of Economic Issues*, 42, 2: 357–65.
- Darity, W. Jr. (1992) "A Model of 'Original Sin': Rise of the West and Lag of the Rest," *American Economic Review*, 82, 2: 162–7.
- Flaschel, P. and Semmler, W. (1988) "On the Integration of Dual and Cross Dual Adjustment Processes in Leontief Systems," *Recherche Economique*, XLII: 403–32.
- Forstater, M. (2004) "Cumulative Causation à la Lowe: Radical Endogeneity, Methodology, and Human Intervention," in G. Argyrous, M. Forstater and G. Mongiovi (eds)

- Growth, Distribution, and Effective Demand: Alternatives to Economic Orthodoxy: Essays in Honor of Edward J. Nell*, Armonk, NY: M. E. Sharpe.
- Heilbroner, R. (1973) "The Paradox of Progress," *Journal of the History of Ideas*, 34, 2: 243–62.
- Hirschman, A. O. (1958) *The Strategy of Economic Development*, New Haven, CT: Yale University Press.
- Kaldor, N. (1967) *Strategic Factors in Economic Development*, Geneva: W. F. Humphrey Press.
- (1972) "The Irrelevance of Equilibrium Economics," *Economic Journal*, 82: 1237–55.
- (1981) "The Role of Increasing Returns, Technical Progress and Cumulative Causation in the Theory of International Trade and Economic Growth," *Economie Appliquee*, 34, 4: 593–617.
- Kurz, H. (1984) "Ricardo and Lowe on Machinery," *Eastern Economic Journal*, 10: 211–29.
- Lowe, A. [1935] (2003) *Economics and Sociology*, London: Routledge.
- (1954) "The Classical Theory of Economic Growth," reprinted in A. Oakley (ed.) *Essays in Political Economics*, Brighton: Wheatsheaf Press [1987].
- (1965) *On Economic Knowledge*, New York: Harper & Row.
- (1976) *The Path of Economic Growth*, Cambridge: Cambridge University Press.
- Marx, K. (1990) *Capital*, Vol. I, London: Penguin Classics.
- Myrdal, G. (1957) *Economic Theory and Under-developed Regions*, London: Gerald Duckworth & Co.
- Nell, E. J. (1988) *Prosperity and Public Spending: Transformational Growth and the Role of Government*, Boston, IL: Unwin Hyman.
- (1992) *Transformational Growth and Effective Demand: Economics after the Capital Critique*, London: Macmillan.
- (1998) *The General Theory of Transformational Growth: Keynes after Sraffa*, Cambridge: Cambridge University Press.
- Pasinetti, L. (1981) *Structural Change and Economic Growth*, Cambridge: Cambridge University Press.
- (1993) *Structural Economic Dynamics*, Cambridge: Cambridge University Press.
- (2007) *Keynes and the Cambridge Keynesians: A Revolution in Economics to be Accomplished*, Cambridge: Cambridge University Press.
- Ricardo, D. (1951) *On the Principles of Political Economy and Taxation*, Vol. 1 of *The Works and Correspondence of David Ricardo*, ed. P. Sraffa, Cambridge: Cambridge University Press.
- Ricoy, C. (1987) "Cumulative Causation," in J. Eatwell, M. Milgate and P. Newman (eds) *The New Palgrave*, London: Macmillan.
- Smith, A. [1776] (1937) *An Inquiry into the Nature and Causes of the Wealth of Nations*, ed. E. Cannan, New York: Modern Library.
- (1986) *The Essential Adam Smith*, ed. R. Heilbroner, New York: W. W. Norton.
- Toner, P. (1999) *Main Currents in Cumulative Causation*, Basingstoke: Macmillan.
- Veblen, T. (1898) "Why is Economics not an Evolutionary Science?," *Quarterly Journal of Economics*, 12: 373–97.
- Walsh, V. (1992) "The Classical Dynamics of Surplus and Accumulation," in J. Halevi, D. Laibman, and E. J. Nell (eds) *Beyond the Steady-state*, London: Macmillan.
- Young, A. (1928) "Increasing Returns and Economic Progress," *Economic Journal*, 38: 527–42.

# 11 Peirce, Veblen, and the introduction of cumulative causation into economic science<sup>1</sup>

*John Hall and Oliver Whybrow*

## Introduction

This chapter traces and also emphasizes strong connections between ideas regarding continuity and continuousness introduced into American philosophical thinking by Charles Sanders Peirce. Some decades after attending a seminar held by Peirce, Thorstein Bunde Veblen introduced these ideas into seminal contributions to social science thinking and economic science.

Connections between Peirce and Veblen, and especially Peirce's influence on Veblen's ideas have been speculated. In his book, *Thorstein Veblen and His America* ([1938] 1972), Joseph Dorfman appears to be the first to note the connections between Peirce and Veblen. Some decades later, Alan Dyer (1986) elaborates on a host of similarities related to scientific inquiry and method between Peirce and Veblen. Dyer (1986: 30–2) stresses that Peirce's seminal contributions to epistemology found their way into Veblen's preference for reasoning by "induction" over "deduction." Dyer (1986: 31) further suggests that Veblen's understanding of, definition of, as well as his use of "deduction" would be more accurately interpreted as a direct borrowing of Peirce's concept of "Abduction."

In addition, Robert Griffen (1998) explores what was initially a short-term contact between Peirce and Veblen at Johns Hopkins University in 1881: a contact that would yield long-term influences on Veblen's thinking. However, Griffen's detailed account of Peirce's influence on Veblen – like Dyer's – remains limited mostly to questions of epistemology, namely what Veblen borrowed from Peirce regarding theory of knowledge and scientific method.

What Dyer and Griffen fail to emphasize – and what we seek to establish in this inquiry – is what we suggest to be Veblen's most important and enduring contribution to economic science. Namely, Veblen sought to lead economic science away from its foundation in Newtonian mechanics, recasting economic science as an evolutionary science. In these efforts, Veblen appears fully indebted to Peirce's contribution to American philosophical thinking, as Veblen relies on concepts advanced by Peirce for developing his understanding of "cumulative causation," and other notions related to processes and changes rooted in continuity and continuousness.



## Peirce on synechism

Charles Peirce devoted his creativity and brilliance toward engaging in numerous areas of inquiry: ranging from geology, to chemistry, to semiotics, to logic, as well as other areas. However, political economy and economic science remained beyond the scope of Peirce's inquiries. Veblen's range of interests was indeed broad – in the tradition of Peirce. Veblen's interests ranged from war and peace to questions of epistemology and even the state of American higher education. Unlike Peirce, Veblen tended to concentrate on and devote the largest portion of his writings to topics related to economic science.

Peirce devoted a portion of his broad inquiry into realms of knowledge toward understanding “continuity” and “continuousness.” Peirce borrowed the term synechism from his reading of ancient Greeks, relying on understandings of *synechismos*, that is, related to *synechés*, suggesting “continuity” or how things are “held together,” as Reynolds (2002: 10–11) teaches us. Following the Greek understanding, Peirce assigned the definition and meaning of “continuous” to the Greek words. Thus, a “synechist,” in Peirce's view, would then be a person who recognizes the importance of continuity and continuousness.

In a philosophical nutshell, synechism appears as a tendency in philosophical inquiry that insists on the necessity of hypotheses involving true continuity. In the 1898 book, *Cambridge Lectures on Reasoning and the Logic of Things*, Peirce ([1898] 1992) teaches us that synechism considers the importance of “Firstness,” “Secondness” and “Thirdness.”

To wit, Firstness is suggested to be wholly related to chance. Secondness would then be characterized as a “brute” reaction to Firstness. Thirdness is then suggested to not be out of relation to Firstness and Secondness. Without Firstness and Secondness, Peirce teaches us, Thirdness “would not have anything upon which to operate” (Peirce [1898] 1992). Peirce's understanding suggests that Thirdness implies an outcome not unrelated to Firstness and Secondness. Hausman (1993: 152–3) suggests that Peirce's Thirdness is wholly unlike Hegel's notion of “synthesis.” Within Hegel's dialectical framework, synthesis is suggested to emerge as a dependent outcome of “thesis” and “anti-thesis.” Peirce rejects the Hegelian deterministic understanding of “synthesis,” and instead insists on the independence of Thirdness from Firstness and Secondness.

Peirce's understanding of continuity and continuousness – as noted by synechism – may be thought to play a fundamental role in philosophical inquiry and to imply broad meanings. So important is continuity that Peirce notes that synechism or the synechist “refuses to believe that when death comes, that the carnal body ceases quickly” (Peirce [1898] 1992). We take this to imply that continuity transcends the meaning and even the significance of bodily death: that bodily death is not really some kind of definitive end in itself. In addition, the synechist fails to distinguish or differentiate between “physical” and “psychical phenomena,” instead suggesting that all phenomena are of one character, with some appearing more material and others more metaphysical.

Joseph Dorfman (1938) notes Veblen to have attended Peirce's lectures when both were at Johns Hopkins University in the early 1880s. With greater respect for detail, Griffen (1998: 733) notes that in the Fall of 1881, Veblen was indeed enrolled at Hopkins and taking Peirce's seminar, "Elementary Logic." Griffen further notes that the topics likely covered in this course included "philosophical questions such as the conception of causation" (Griffen 1998: 733).

Dyer also notes that Veblen did attend Peirce's lectures when both were at Hopkins. In addition, Dyer suggests that Veblen's failure to directly reference Peirce is not to be taken that Veblen was not borrowing from Peirce. Dyer (1986: 30) notes that in Veblen's article "Kant's Critique of Judgment" Veblen fully understood Peirce's concept of "Abduction." In this vein of thinking, we would also like to speculate that Veblen not only knew of and understood Peirce's concept of synechism, but the concept and principles of synechism and continuousness would later emerge as the most seminal understanding of "cumulative causation," an assumption and process at the core Veblen's understanding of social and also economic processes, that served to lay a foundation for his understanding of social and cultural evolution. Where Veblen extends Peirce is that in his understanding of continuity and continuousness he also strongly implies and even emphasizes "connection" and "connectedness," what Peirce implies but fails to ostensibly emphasize.

We think it is appropriate to re-emphasize that Veblen was well schooled in philosophy, completing his Ph.D. in this discipline at Yale University in 1884. As Mark Blaug (1986: 258) teaches us, Veblen went on to study for and earn a subsequent Ph.D. in economics at Cornell University. A large part of the richness in Veblen's contribution to economic science, and one of the reasons why we still read Veblen so avidly and grapple with the seriousness of his ideas, is that he brought to the economics discipline a profound as well as a thoroughly schooled knowledge of philosophical inquiry. His knowledge of philosophy proved particularly effective in his rethinking, reformulating, and challenging neoclassical economics, especially that tradition represented by his contemporary, Irving Fisher.

Nicholas Georgescu-Roegen (1998: 387) stresses that Fisher, especially, relied on a mechanistic approach: growing out of what he terms "classical mechanics." Schooled in philosophy, Veblen effectively countered Fisher's elementary and mechanical understanding of economic processes through introducing a Peircian influenced approach to economic science, an approach emphasizing continuity and continuousness as it relates to an interplay between and among material and immaterial elements and forces in the creation of social and economic processes.

### **Peirce's influences in Veblen**

In his "Instinct of Workmanship," Veblen concerns himself with ways in which material and immaterial changes come about and as well engender further changes in the material and immaterial. Veblen can be quoted:

The ways and means, *material and immaterial*, by which the native proclivities work out their ends, therefore, are forever in a process of change, being conditioned by the changes *cumulatively* going forward in the institutional fabric of habitual elements that govern the scheme of life.

(Veblen [1898b] 1993: 185; emphases in original)

What Veblen is suggesting is that elements, both material and immaterial, or, as Peirce asserts, physical and psychical, are characterized by continuousness and also connectedness. Although Veblen emphasizes “change” more than does Peirce, when doing so he suggests that change or evolution in society and economy are also integral to continuousness and connectedness, as change and evolution are engendered in the interplay of the material and immaterial.

In his article “Why is Economics Not an Evolutionary Science?,” first appearing in the *Quarterly Journal of Economics* in 1898, Veblen stresses advances made in the natural, and especially biological sciences as these areas of inquiry moved away from taxonomy and toward an evolutionary approach to these disciplines. Biologists, in Veblen’s ([1898a] 1993: 131) view, shifted inquiry away from what he terms the “taxonic structures of reefs” and toward the living organism, the polyp – as it were – whereby the living habits of this modest organism engender changes and evolutionary processes found in reefs, while also responding to changes in reefs. To quote Veblen: “[a]ll the talk about cytoplasm, centrosomes, and karyokinetic process means that the inquiry now looks consistently to the life process, and aims to explain it in terms of *cumulative causation*” (Veblen [1898a] 1993: 136; emphasis in original).

Shifting his inquiry away from advances in biology, but relating these advances to human beings as well as to social science’s ability to deal with continuousness, continuity, and also evolution, Veblen notes:

The economic life history of the individual is a *cumulative process of adaptation* of means to ends that *cumulatively change* as the process goes on, both the agent and his environment being at any point the outcome of the last process. His methods of life today are enforced upon him by his habits of life carried over from yesterday and by circumstances left as the mechanical residue of the life of yesterday.

(Veblen [1898a] 1993: 139; emphases in original)

Veblen adds:

What is true of the individual in this respect is true of the group in which he lives. All economic change is a change in the economic community – a change in the community’s methods of turning material things to account. The change is always in the last resort a change in habits of thought (Peirce’s psychical). A given contrivance for effecting certain material ends becomes a circumstance which affects the further growth of habits of thought – habitual methods of procedure – and so becomes a point of

departure for further development of methods of compassing the ends sought and for the further variation of ends that are sought to be compassed.  
(Veblen [1898a] 1993: 139–40)

To Veblen, human and societal activities are far from being pointless, random, and without tendency. Human beings and society exhibit – if not a predetermined or even a specified direction – at least a tendency. Veblen notes that:

[E]conomic action is teleological, in the sense that men always and everywhere seek to do something. [Veblen adds] It is necessarily the aim of such an economics to trace the *cumulative* working-out of the economic interest in the cultural sequence.

(Veblen [1898a] 1993: 140–1; emphases in original)

What Veblen poses as continuity, continuousness, connection, and connectedness – as well as the profound link between the material and immaterial – may be seen not only as drawing heavily from Peirce’s thinking on synechism, but also as an extension of its basic tenets. Veblen, just like his Hopkins’ professor, Charles S. Peirce, understands that evolutionary change is integral to the interaction of the material and immaterial. In this way, Peirce’s seminal contribution is advanced forward by Veblen – away from pure philosophy – and into the realm of economic and social inquiry. Veblen was writing on a diverse range of subjects in economic inquiry in the final decades of the nineteenth century and in the first three decades of the twentieth century, moving forward with a Peircian approach.

## Conclusion

This chapter has sought to establish that continuity and continuousness can be found in the chain of ideas linking Charles Peirce’s understanding of synechism to Thorstein Veblen’s understanding of “cumulative causation.” Our research findings suggest that cumulative causation emerged as a key term, concept, and understanding used in Veblen’s inquiry into the dynamic interplay of the material and immaterial in economic and social processes, and thus serves as the foundation for his efforts to develop an evolutionary theory for economic science. Phrased differently, Peirce provided a foundation on which Veblen could build his understanding of process, change, and social and cultural evolution. Veblen understood from Peirce’s teachings that institutional change and social and cultural evolution were engendered not only through changes in the material realm, but also through dynamic and explainable interactions between the material and the immaterial.

Both Peirce and Veblen were exposed to and challenged by contributions of Charles Darwin, and especially Darwin’s hypothesis regarding the central role of natural selection, what later developed into a theory of biological evolution. As Goudge (1964: 323) teaches us, Peirce – influenced by Darwin’s advances –

attempted to develop an evolutionary philosophy. Veblen, in the tradition of Professor Peirce, as well as Darwin, sought to transform economics into an evolutionary science.

While Veblen's personal contacts with Professor Peirce appear limited to just a few months back in Baltimore in 1881, the exposure nevertheless appears to have engendered powerful influences. We believe these influences are aptly characterized as generating continuity and continuousness, as well as displaying connection and connectedness in the ideas of Peirce and Veblen: thus serving as a useful example of cumulative causation.

## Note

- 1 This chapter is based on a paper presented at the annual meeting of the Association for Evolutionary Economics in New Orleans, LA, January 4–6, 2008.

## References

- Blaug, M. (1986) *Great Economics Before Keynes*, London: Cambridge University Press.
- Dorfman, J. (1938, 1943; 7th edn 1972), *Thorstein Veblen and His America*, Clifton, NJ: A. M. Kelley.
- Dyer, A. (1986) "Veblen on Scientific Creativity: The Influence of Charles S. Peirce," *Journal of Economic Issues*, 20, 1: 21–41.
- Georgescu-Roegen, N. (1998) "The Entropy Law and the Economic Process," in G. Hodgson (ed.) *The Foundations of Evolutionary Economics 1809–1973, Vol. II*, Northampton: Edward Elgar.
- Goudge, T. (1964) "Peirce's Evolutionism – After Half a Century," in E. Moore and R. Robin (eds) *Studies in the Philosophy of Charles Sanders Peirce*, Amherst, MA: The University of Massachusetts Press.
- Griffen, R. (1998) "What Veblen's Theory Owed to Peirce – The Social Theory of Logic," *Journal of Economic Issues*, 32, 3: 733–57.
- Hausman, C. (1993) *Charles S. Peirce's Evolutionary Philosophy*, New York: Cambridge University Press.
- Peirce, C. S. ([1898] 1992) *Cambridge Lectures on Reasoning and the Logic of Things: The Cambridge Conferences Lectures of 1898*, ed. K. L. Ketner, Cambridge, MA: Harvard University Press.
- Reynolds, A. (2002) *Peirce's Scientific Metaphysics: The Philosophy of Chance, Law, and Evolution*, Nashville, TN: Vanderbilt University Press.
- Veblen, T. ([1898a] 1993) "Why is Economics Not an Evolutionary Science?," in R. Tilman (ed.) *A Veblen Treasury: From Leisure Class to War, Peace, and Capitalism*, Armonk, NY: M. E. Sharpe.
- ([1898b] 1993) "The Instinct of Workmanship," in R. Tilman (ed.) *A Veblen Treasury: From Leisure Class to War, Peace, and Capitalism*, Armonk, NY: M. E. Sharpe.

# 12 Veblen's cumulative causation and the origins of money in Mesopotamia

*Alla Semenova*

## Introduction

The aim of this chapter is to improve the Chartalist perspective on the origins of money by applying it to the socio-economic modes of organization of ancient Mesopotamia in the fourth millennium BC. It will be argued that the Chartalist account does not yet adequately address the prevalent institutions and transactional modes of ancient societies. To bring Chartalism into congruency with the institutional context of ancient societies, I apply Veblen's theory of cumulative causation to the factual evidence of Mesopotamian clay tokens and their assemblages into clay cases known as "bullae." After all, laying out cumulative causation (CC) theory in "Why is Economics Not An Evolutionary Science," Veblen referred to the eminent anthropologist M. G. de Lapouge: "Anthropology is destined to revolutionize the political and the social sciences as radically as bacteriology has revolutionized the science of medicine" (Veblen 1898: 373).

The Chartalist account considers the introduction of money as a means of forcing a population into a debt relationship to a central public authority. This is precisely the point where CC can enhance Chartalism. Payment obligations (or a debt relationship of a population to a central public authority) in the form of goods and services had been in place for a long time before monetization was introduced. Thus, CC perceives money as having cumulatively emerged to *better control* the fulfillment of payment obligations to a central public authority that were *already in effect* (or, equivalently, reducing their avoidance). This hypothesis is especially warranted given the absence of bureaucratic tools, such as writing,<sup>1</sup> census systems and formal means of personal identification that would allow to detect those members of an ever-growing and impersonal community that had payment obligations to a central public authority.

## The Chartalist theory of the origins of money

The Chartalist (or state money) framework is rooted in the seminal contribution of G. F. Knapp's book, *The State Theory of Money* ([1905] 1924). A key distinction of the Chartalist approach to the nature and origins of money is its alternative

to the Metallist (Mengerian) approach, i.e. the "exchange theory of money." According to the latter, money emerged as a medium of exchange embodied in a valuable commodity with optimal characteristics for the purposes of exchange (such as divisibility, portability, durability, homogeneity), namely a precious metal (Menger 1892). K. Polanyi ([1944] 1962, 1957a, 1957b) and the Substantivist School of anthropology demonstrated that the exchange-based approach, which projects a pre-eminence of (market) exchange relations into ancient societies and primitive communities, is not supported by historical and anthropological evidence, and is reliant upon a false characterization of human "nature" as *homo economicus*. The Substantivist position is shared by the theories of money developed by B. Laum (1924), P. Grierson (1975, 1977), R. Seaford (2004) and others who pointed out the marginality of commercial relations, both foreign and domestic, in archaic and ancient societies of the Mediterranean when the earliest Western money emerged.

Countering the Metallist perspective, the Chartalist framework denies the origins of money in a context of market exchange. As an alternative, Chartalism views the origins of money in a debt relationship of the population to a central public authority (e.g. a state, a monarchy).<sup>2</sup> Rather than a medium of exchange invented by individuals trading with each other, money is introduced by the state (or any other central public authority) as a unit of account in which debts to the state are denominated and have to be repaid. Rather than a commodity with an intrinsic value, money is, first and foremost, a state's "token" (or a "ticket") that acquires its value from its power to extinguish debt obligations to the state. The need on the part of the population to pay their debts induces them to supply goods and labor services to the state in exchange for that which is necessary to repay their debt obligations (i.e. money) (Wray 1998: 54). The functions of money as a means of payment and a medium of exchange in private transactions are derived from its primary role as a unit in which debt obligations to the state are denominated and must be paid (ibid.: 23–9, 37, 51; Goodhart [1998] 2003: 5–9; Knapp [1905] 1924: 25–53).

So far there are two main mechanisms of monetization outlined by Chartalism. In the first mechanism, the state introduces money by means of imposing debts on the population ("tax" obligations to use *modern* terminology). Money and taxes are used "as a means of *inducing* the population to supply goods and services to the state, supplying in return the money that will be used to retire the tax liability" (Wray 1998: 37; emphasis added). In this, money is an alternative to using overt coercion (i.e. overt force, violence) to extract resources from the local population. Although it is an alternative to overt force, money performs a coercive function as it enables the state to extract real resources and services from its population. In the absence of money and taxation, it is argued, the state would not be able to obtain anything from its population except through a threat of violence.<sup>3</sup>

According to the second Chartalist mechanism of monetization, money may have been introduced by the state as a means of facilitating its "fiscal base," *after* a debt relationship of the population to the state *is already in place*. This



argument is rooted in a proposition that collecting and managing the supplies of various goods obtained as in-kind payments from the population could become quite burdensome and inconvenient. In addition, other problems associated with in-kind payment of taxes and tribute are identified by Chartalists and linked causally to the origin of state money. For example, taxation in-kind could produce a fiscal base imbalance, i.e. an overabundance of some types of goods and a shortage of others (Tymoigne and Wray 2005: 2; Goodhart [1998] 2003: 7). When taxes are received in goods or labour, the balance of goods (and labour) obtained will not be that required for public sector expenditures" (Goodhart [1998] 2003: 7). Moreover, in-kind taxation could create an "incentive for the taxpayer to provide the lowest quality goods" in payment of taxes (Tymoigne and Wray 2005: 2). To eliminate such problems, it was desirable for the state to introduce a mechanism through which payments could be standardized (ibid.: 2; Wray 1998: 50–1). Rather than stipulating the specific types, qualities and quantities of goods that had to be paid as tribute or taxes, the state "could insist on payment in units [...] it had itself created" (Kraay 1964: 90). Hence, in the second Chartalist mechanism of monetization, money was invented by a central public authority as a means of standardizing tribute or taxes levied upon the population. Money emerged as a practical solution to the problems of fiscal base cumbersomeness, imbalance, and quality inferiority. As in the previous mechanism of monetization money is not a commodity with an intrinsic value, but a "token" (or "ticket") whose value is derived from its power to extinguish debt obligations to a central public authority. The use of money as a means of payment and a medium of exchange in private transactions it derived from its primary role as a unit of account in which debt obligations are denominated and have to be paid. Discussing the introduction of money in ancient societies, Wray summarizes the second Chartalist mechanism as follows:

Palaces created the money units to standardize payment of taxes. Use of money in private transactions derived from tax debts [...] Once a money tax was levied on a village and later on individuals, the palace would be able to obtain goods and services by issuing its own money-denominated debt.

(Wray 1998: 51)

Thus, upon collection of money via taxation, the central public authority could choose the type, quality, and quantity of goods to purchase, as well as decide when to purchase them (Wray 1998: 50). The population would be willing to accept the otherwise worthless (i.e. intrinsically worthless) money in payment for goods and services purchased by the state because a universal demand for "that which is necessary to pay taxes" would be created (ibid.: 36–7). Note that the second Chartalist mechanism of money's introduction relies upon a trade relation between a state and its citizens. While commercial (market) relations between states and citizens (population) are characteristic of private property regimes or modern colonial regimes, they were not typical for ancient despotic-communal societies characterized by possession regimes.<sup>4</sup>

In sum, the Chartalist framework offers two possible mechanisms through which an economy could be monetized. The first mechanism takes place in a context of a previously non-existent debt relationship of a population to a sovereign power. Direct, once-and-for-all monetization of an economy is a means of coercing (though indirectly and without recourse to overt force/violence) a local population into supplying goods and services to a central public authority. In this, monetization could be viewed as a solution to the work effort inducement problem. The second mechanism presupposes that the payment of taxes and tribute (i.e. a debt relationship of a population to a central public authority) is already in place. Money is introduced as a practical solution to the problems of fiscal base cumbersomeness, imbalance, and quality inferiority.

### **Veblen's cumulative causation: "institutionalizing" Chartalism**

Against this background I argue that both Chartalist mechanisms of monetization could not have been introduced *ex nihilo*. Rather, monetization had to *evolve* through a Veblenian process of cumulative causation adapting means to ends in a concrete historical process.

[E]volutionary economics must be the theory of a process of cultural growth as determined by the economic interest, a theory of a cumulative sequence of economic institutions stated in terms of the process itself.

(Veblen 1898: 393)

Any evolutionary science [...] is a closeknit body of theory. It is a theory of a process, of an unfolding sequence.

(ibid.: 375)

As Veblen argued, institutional conventions undergo a "cumulative process of development" (Veblen 1901: 83) or "cumulative change" (Veblen 1898: 387). Veblen understood institutions as "a cumulative process of adaptation of means to ends" (ibid.: 391), as a "cumulatively unfolding process or an institutional adaptation to cumulatively unfolding exigencies" (Veblen 1900: 264).

Applying Veblen's notion of cumulative change in means–ends relationships to the origins of money in ancient Mesopotamia, I argue that the initial *end* of monetization was to maintain or improve control over the fulfillment of *existing* payment obligations in-kind. The *means* had to be adapted to a specific institutional environment and the problems (exigencies) posed by it.

To figure out the institutional context and exigencies of different societies, Veblen urged the need to carefully examine historical materials:

The scrutiny of historical details serves this end by defining the scope and character of the several factors causally at work in the growth of culture, and, what is of more immediate consequence, as they are at work in the

shaping of the economic activities and the economic aims of men engaged in this unfolding cultural process as it lies before the investigator in the existing situation.

(Veblen 1901: 80)

[A]ny science, such as economics, which has to do with human conduct, becomes a genetic inquiry into the human scheme of life; and where, as in economics, the subject of inquiry is the conduct of man in his dealings with the material means of life, the science is necessarily an inquiry into the life-history of material civilization, on a more or less extended or restricted plan.

(Veblen 1909: 627–8)

As far as the institution of money is concerned, Veblen noted the tendency to project the scholar's ideal of money into ancient history, rather than inquire into historical causal relations of its origins and development: "Money is [...] discussed in terms of the end which, [...] it should work out according to the given writer's ideal of economic life, rather than in terms of causal relation." (Veblen 1898: 383). Veblen hoped that historical anthropological inquiry would alter the then (and now) prevailing trend of projecting modern institutions and habits of thought into ancient history. The German Historical School and the Substantivist School of anthropology in the tradition of K. Polanyi also embarked on this task of a detailed historical and anthropological inquiry to demonstrate that "human beings [...] do not always live under the same [...] social institutions" (Schmoller [1949] 1963: 227). As far as the institution of money is concerned, the implication is that its origins, evolution, and social functions must be closely linked to the particular modes of transactions and socio-economic integration of different societies (Dalton 1971: 169). Therefore, it would be futile to project the functions performed by money today or in recent history (e.g. means by which *monetary* taxes are paid) into ancient societies characterized by distinct institutions and modes of socio-economic integration not found in modern history. As Dalton (1971) put it:

The essential point is that the characteristics of money in any economy, including our own, express the basic organization of that economy [...] Money is linked to specific modes of transactions.

(Dalton 1971: 169)

Nevertheless, both the exchange-based and the Chartalist perspectives on money commit this fallacy of projecting modern institutions, habits of thought, transactional modes, and forms of socio-economic integration into ancient history. Whereas the exchange-based approach projects a notion of a pre-eminence of market exchange into ancient and even primitive forms of socio-economic organization, the Chartalist perspective assumes ancient societies to exhibit a modern-type state entity (though it may be referred to as a palace, a ruler, a governor) and modern forms of socio-economic organization (such as a *monetary* taxation; a trade relationship between a state and its citizens).

Of course, this is not to imply that some of the functions performed by money today and some of its modern features could not have been present in the earliest stages of monetary evolution. The issue is with assuming these functions and features without a proper historical verification of their plausibility. In particular, it should be verified whether money's primary function as a medium of exchange (Metallism) or its primary role as a unit in which taxes are denominated and have to be repaid (Chartalism) would be compatible with the transactional modes, forms of socio-economic organization, and the prevalent institutions of ancient societies.

### **The institutional context of ancient Mesopotamia**

An inquiry into ancient Mesopotamian in the fourth millennium BC demonstrates that the transactional modes and forms of socio-economic integration were not characterized by a prevalence of market exchange, *monetary* taxation, or a trade relationship between a central public authority and its population. Rather, a socio-economic structure of such societies was characterized by a centralized collection and redistribution of goods, as well as a complete infiltration of civic, economic, and religious matters. For example, the prevalent social institution – a Mesopotamian temple occupying a position of economic pre-eminence – collected the agricultural surplus from its local villages and further redistributed it among the members of the temple priest-kingship, priesthood, and their supporting apparatus (i.e. various officials, soldiers, professional experts, hired laborers, slaves). Not only distribution, but production cycles as well, were administered from the center by a temple's bureaucratic personnel (Oppenheim 1956: 31–5). Schmandt-Besserat presents a similar picture of ancient Mesopotamian economies outlining their three major institutional components:

- 1 a temple which conferred meaning and pomp on the act of giving;
- 2 an elite who administered the communal property; and
- 3 commoners who produced surplus goods and surrendered them to the temple.

(Schmandt-Besserat 1996: 107)

Thus, the factual evidence of ancient Mesopotamian societies speaks against the Chartalist mechanisms of money's introduction. First, a debt relationship of a population to a central public authority (a temple) was firmly in place before monetization was introduced, so money could not have come into being as a means of coercing (though without recourse to direct violence) a local population into supplying goods and services to a public authority. Second, payment obligations to a central public authority were denominated and had to be settled in goods and services, rather than in monetary units such as tokens, tickets, coins, and so on. Third, Mesopotamian temples had a coercive system to enforce a collection of in-kind payments (Schmandt-Besserat 1996: 108), so that money could not have emerged as a solution to a work-effort inducement problem.

Given this knowledge, I will attempt to develop a Veblenian, historically informed mechanism of monetization that could have taken place in an institutional context of ancient Mesopotamia in the fourth millennium BC. Such a mechanism will be grounded in the actual transactional modes and forms of socio-economic integration of ancient Mesopotamia rather than in preconceptions based on modern institutions and habits of thought. I will hypothesize how a conception of money could have evolved out of practical solutions to the material problem situations faced by ancient Mesopotamian priest-kings and their bureaucratic apparatus. I use the evidence of ancient Mesopotamian clay tokens and their assemblages into clay cases known as “bullae.” The interest in these artifacts is related to the Chartalist conception of money as a state’s “pay-token” devoid of *intrinsic* value (i.e. the value of a material from which it is made). However, the goal must not be to project a modern Chartalist conception of money as a state’s “pay-token” into ancient history. Rather, the aim is to conduct a historical test whether a concept of a state’s “pay-token” could have emerged in ancient societies, and if so, *how*. As was noted above, the issue under consideration is a projection of modern institutions into ancient history *without a historical verification of their plausibility* under the prevalent institutions, transactional modes, forms of socio-economic integration, and so on specific to ancient societies.

### Clay-tokens and bullae: the conventional interpretation

Commonly considered among the earliest counting and inventory-keeping devices, Mesopotamian clay-tokens (c. 5000–2500 BC) are small geometrically (as well as naturalistically) shaped objects. With each token representing a counted unit<sup>5</sup> as indicated by a token shape (and, later, markings on its surface), a collection of tokens would represent a sum of the units counted. Such a token assemblage could be “put away in a safe place with restricted access” in one of the earliest human attempts at inventory-keeping (Nissen *et al.* 1993: 11). A one-to-one correspondence between a token and a unit counted, as well as a variety of token shapes (and, later, surface markings), made this inventory-keeping system cumbersome and awkward (Schmandt-Besserat 1996: 15–16, 95–6; 1981: 283).

While the earliest Mesopotamian tokens were characterized by a plain surface (the so called “plain” tokens), somewhere in the middle of the fourth millennium BC various markings and inscriptions appeared on the tokens’ surfaces (the so-called “complex” tokens). This evolution in token design was paralleled by the process of urbanization, increased specialization of labor, and proliferation of manufactured goods. While “plain” tokens represented agricultural staples (as well as cattle, land, and so on), “complex” tokens became a method to signify city merchandise (Schmandt-Besserat 1996: 16, 82). Whereas plain tokens were a product of rank societies, “it was the advent of the state which was responsible for the phenomenon of complex tokens” (*ibid.*: 107).

Although the majority of tokens found by archeologists come from public districts (e.g. ancient Mesopotamian temples, sacred precincts, palaces, ware-

houses), token assemblages were found in private (secular) quarters as well (e.g. Schmandt-Besserat 1996; Jasim and Oates 1986; Nilhamn 2002). However, as Nissen cautions, this distribution of tokens should not be attributed with universal significance considering that public areas have been excavated more often than private quarters (Nissen 1988). When found in domestic settings, small quantities of tokens were commonly stored in kitchen utensils such as jars and pots (Schmandt-Besserat 1996: 37).

The evolution of the inventory-keeping system from “plain” to “complex” tokens was also accompanied by changes in the methods of storing them. With such innovations as a perforation in the middle of a token surface (c.3500 BC) tokens could be strung together – presumably to separate tokens dealing with the same “transaction.” Another innovation was to enclose tokens into spherical clay cases known as bullae, the outer surface of which was sealed with official seal impressions (from c.3700–3500 BC to c.2600 BC). Apart from separating tokens dealing with the same “transaction,” a sealed bulla had the advantage of preventing unauthorized access to the tokens enclosed. At the same time, the sealing of a bulla allowed the identification of the official parties to the transaction (Schmandt-Besserat 1996: 39–40, 44).<sup>6</sup>

Sometimes, an outer surface of a clay bulla could bear the impressions of the tokens it contained. This innovation was crucial for the evolution of writing, since it was soon realized that the tokens enclosed within a clay bulla were made redundant by the presence of tokens’ impressions on its outer surface. Consequently, clay tablets bearing impressed markings in the shape of tokens came into being (c.3500–3100 BC). This invention further led to the take-off of the Sumerian pictographic script (c.3100–3000 BC) (Schmandt-Besserat 1996: 125).

The ancient historians’ and historical anthropologists’ primary interest in clay tokens and bullae is related to their function as one of the earliest accounting and inventory-keeping devices, and, even more so, to their significance in the evolution of Sumerian writing. The scholars are not so much concerned with exploring any possible role that the system of tokens and bullae may have played in the enforcement of in-kind payments of “taxes” and tribute to the ancient Mesopotamian temples. While some historians have hinted at such a role (e.g. Schmandt-Besserat 1981: 982; 1982: 875–6; 1996: 108), they failed to formulate a concrete mechanism that would *explain how* the system of tokens could play a part in the enforcement of in-kind “taxes” and tribute owed to Mesopotamian temples.

Notably, in one of her earlier publications, Schmandt-Besserat seriously questioned the function of tokens as counters or inventory-keeping devices, saying that “[t]his idea of keeping track of individual food reserves is not fully convincing” (Schmandt-Besserat 1982: 874). Arguing that the token system suggested a “large-scale constraint rather than the invention of a household gadget” (ibid.: 875) she suspected that the fourth-millennium BC tokens were directly linked to involuntary payment of obligations to a public authority:

[...] tokens of the 4th millennium BC may be viewed as records of the pooling of resources by means of ceremonial ritual. [...] the offerings were

mandatory rather than voluntary and can be regarded as taxation. The first use of writing and, for that matter, the last use of tokens, was [...] a means of control upon the delivery of goods and ultimately a control on the production of real goods. Can it be inferred that a system of redistribution through ceremonial ritual had its origin earlier and the tokens were related to such a process?

(ibid.: 875)

While in the paragraph above, Schmandt-Besserat (1982) recognized the function of tokens as “records of the pooling of resources” through a process of taxation that had its origins in ceremonial rituals, and characterized tokens as “a means of control upon the delivery of goods,” she failed to explain *how* and *via what mechanism* tokens could perform such instrumental functions. Similarly, whereas Schmandt-Besserat (1996) suggested that “complex tokens played a part in the collection of taxes and tribute that is typical of a state economy” (Schmandt-Besserat 1996: 108), she fell short of providing a concrete mechanism that would explain *how* the complex tokens could play such a role. Likewise, while arguing that the invention of complex tokens and bullae reflected “the need for enforcement of deliveries of goods owed to the temple” (ibid.: 110), Schmandt-Besserat did not specify a mechanism that would describe *how* the tokens and bullae could enforce such deliveries.

Formulating a mechanism that could explain how the system of tokens played an instrumental role in the process of in-kind “taxation,” the following section provides a reformulation of the Chartalist account of monetization based upon Veblen’s notion of cumulative causation in means–ends problem solving. This account will be specific to the institutional context of ancient Mesopotamia, while it will preserve the Chartalist conception of money as a token issued by a central public authority in a context of “taxation.” However, *non-monetary* “taxation” is a point of departure for formulating the alternative monetization mechanism, while the central public institution concerned is that of an ancient Mesopotamian temple. The conception of temple-issued money specific to ancient Mesopotamian circumstances will differ from the conventional Chartalist conceptualization of money as a unit of account in which monetary taxes are denominated and have to be paid because it will emerge from a historical analysis rather than from a projection of modern habits of thought into ancient history.

### **Clay-tokens and bullae as “certificates of contribution”: a CC hypothesis of the origins of money**

Building on Schmandt-Besserat’s (1996) interpretation of the fourth-millennium clay tablets as the “official receipts of commodities delivered by individuals or guilds” and stored in the public archives, (e.g. temple premises, sacred precincts) (Schmandt-Besserat 1996: 105) this paper suggests that “individuals or guilds” acquired some form of a “receipt” as well, signifying their *fulfilled* contribution to the public sector (i.e. a temple). Because it is debatable whether *individuals* or



*collective entities* such as families or villages were responsible for in-kind payments to Mesopotamian temples in the fourth millennium BC, the terminology of a "taxation unit" is adopted here.

Clay-tokens and their assemblages into bullae could have served as official receipts handed to the compliant "taxation units" upon delivery of goods and services to the temple. In this way, a token "receipt" would serve as a certificate of a fulfilled contribution to a temple (or, simply, a "certificate of contribution"). Such "certificates" could be issued to compliant "taxation units" with an aim of retrieving and punishing non-compliant parties. More specifically, when the temple personnel would be commissioned to the local villages, charged with a task of verifying contributions, compliant "taxation units" would have no difficulty presenting their "certificates" of fulfilled contributions (i.e. tokens), while the non-compliant parties (i.e. those without a token "certificate") could be easily identified. A "certificate of contribution" (i.e. a token) presented to the temple personnel would be immediately confiscated to prevent its possible transfer to a non-compliant "taxation unit." Charged with a task of collecting tokens from the villagers, these temple personnel could be referred to as "public collectors." The villagers who failed to present token "certificates" to the public collectors could be punished in various ways, including being channelled into a pool of public laborers. The clay tablet "receipts" stored at the public archives could be compared against the token "receipts" collected from the villages in an attempt to identify any possible counterfeit (of tokens) on the part of the local population. Upon this procedure, tokens could be discarded as the "transactions" they represented had been concluded.

Why would such a mechanism be in place? The key is that in the fourth millennium BC Mesopotamia, census systems were underdeveloped or not yet in place (Modelski 1997), while any formal means of personal identification of the villagers were absent. If the temple bureaucracy could record the compliant "taxation units" on clay tablets, how could they know about *all potential* "taxation units"? The problem posed by this situation is that "tax-evading" parties could not be easily identified. Moreover, even if, hypothetically, the non-compliant parties could be identified it would be difficult if not impossible for the temple personnel to retrieve them given that the villagers did not bear any means of personal identification<sup>7</sup> and their places of residence were most likely unknown to the temple personnel. Note that the fourth-millennium BC Mesopotamia represented a highly developed urban civilization. To appreciate why the temple officials could not keep a memory track of all potential payers of "taxes" and tribute (i.e. their names, dwelling places) let us consider some population estimates of the ancient Mesopotamian cities. For example, while it is estimated that the City of Eridu hosted from 6,200 to 10,000 people *c.*3700 BC, the population of Uruk was projected at 14,000, 20,000 and 50,000 people *c.*3700 BC, 3400 BC, and 3100 BC, respectively (Modelski 1997, Table 2). Notably, population census systems were not yet in place: "censuses are nonexistent, even though the practice of counting e.g. armies and battle casualties does seem to be taking hold by the end of the period [fourth millennium BC]" (ibid.).

Of course, the above hypothesis, describing ancient Mesopotamian tokens as “certificates” of fulfilled contributions to a temple, raises a number of issues. To begin with, this hypothesis relies upon an assumption of non-compliance, i.e. an idea that at least some portion of the population did not *voluntarily* pay “taxes” and tribute to the temple. Indeed, such an assumption may be considered as a projection of modern habits of thought into ancient societies. If “taxes” and tribute in these societies were part of *voluntary* religious practices, in which religious leadership “conferred meaning [...] to the act of giving” (Schmandt-Besserat 1996: 107), personifying temple contributions as “gifts for the gods” (ibid.: 105), then tokens could be viewed as “counter-gifts” bestowed by a temple upon a contributing community in a social context of “reciprocity.” Here, an act of giving was “in an essential sense always the first half of a reciprocal action, the other half of which was a counter-gift” (Finley 1965: 62). Thus, the temple could “reciprocate” (“on behalf of god”) by issuing counter-gifts in the form of tokens and bullae, as well as by promising security and protection to its community<sup>8</sup> (Semenova 2008). As Polanyi (1957a) emphasized, “reciprocity demands adequacy of response, not mathematical equality” of the goods or services reciprocated (Polanyi 1957a: 73). Besides, the temple and the artifacts it issued could serve as the salient attributes of the villagers’ collective identity, thus providing a basis for social cohesion. Therefore, it must be further explored how and why non-compliance evolved out of this voluntary and “reciprocal” religious context. Given space limitations, this issue cannot be dealt with here.

The nature of “taxation units” is another issue posed by the hypothesis developed in this chapter. Were they individuals or collective entities, such as a family, a clan, or a village? While Schmandt-Besserat (1996) has argued that “taxes and tribute” were paid by “individuals and guilds” (Schmandt-Besserat 1996: 105, 108), Oppenheim (1956) dismissed the notion of an *individual* “taxpayer,” arguing, instead, that the ancient Mesopotamian “villages contained a number of families [...] *paying taxes collectively* [...]” (Oppenheim 1956: 35; emphases added). Such an “obligation to pay taxes collectively” – argued Oppenheim – “counteracted individual deflections” (ibid.: 35), pointing to the strength of communal bonds among the villagers (ibid.: 30).

If payments were made collectively, as Oppenheim (1956) has argued, then could a “tax receipt” or a “certificate of contribution” of a whole village community be embodied in a clay bulla? A unified whole of a bulla “certificate” as opposed to a loose and detached token “receipt” would signify the unity of a contributing community (such as a village) and their collective effort in meeting a payment obligation. In such a case, the symbolic wholeness of a clay bulla would most likely be preserved, i.e. a bulla would never be opened (broken) to reveal and redistribute its token components. In fact, most of the clay bullae unearthed by archeologists were found intact (Jasim and Oates 1986: 350). Thus, it is less likely that a clay bulla could be broken and the tokens enclosed redistributed among the villagers. On the other hand, the quantities handled in clay bullae were typically small: on average, an equivalent of about five bushels of grain or five *sila*<sup>9</sup> of oil (Schmandt-Besserat 1996: 103), which is more

suggestive that "taxes" and tribute were paid by small households or small production guilds. No one knows for sure.

Another possibility is to suppose that the amounts of tax contributions increased gradually, requiring temple bureaucracy to issue more tokens (i.e. "receipts") to the contributing parties. Note that a token stood in a one-to-one correspondence to a unit counted (whether a discrete entity or a specified measure of a certain product) (Schmandt-Besserat 1996). To prevent the loss of "receipts" as their numbers increased, they may have been enclosed into bullae. Alternatively, bullae could serve as a response to the problem of counterfeit: villagers could eventually attempt to produce their own tokens if access to clay was not restricted by temple officials. An enclosure of token "receipts" into an officially sealed bulla would provide for their authenticity.

It is not excluded that by the end of the fourth millennium BC, tokens could eventually serve as a means of payment to public laborers, signifying their contributions to the socio-economic provisioning process. Public laborers were typically supported by daily disbursements of grain (as well as beer), where a daily grain ration corresponded to the amount held in a bevel-rim bowl (Nissen 1988: 93–4). By the close of the fourth millennium BC, however, these bevel-rim bowls, mass produced in millions, disappeared from circulation (ibid.: 84). Tokens could have been used afterwards as an alternative form of payment to public laborers superseding the daily ration disbursements in bevel-rim bowls. While a system of in-kind payments to the public workforce involved all the inconvenience associated with collection, storage, transportation, management, and redistribution of food staples, administering labor payments in the form of tokens would have been much easier. A public laborer, in turn, could regularly exchange his tokens (i.e. his "certificates of contribution") for subsistence in a nearby village. Of course, this mechanism would only take place in those circumstances where public works projects were located in the vicinity of a village. Note that with a universal demand for tokens created by the need to present them to a public collector, the villagers would eagerly accept the tokens from the public laborers. Such villagers would no longer be required to make an actual in-kind payment to the temple, as their contribution would be fulfilled by providing the means of subsistence to the public workers. Lending support to this hypothesis, archeologists and ancient historians believe that tokens representing labor were actually in use. Shaped in the form of a tetrahedron they are perceived as representing two different units of labor such as one day's or one week's work. Alternatively, tetrahedron-shaped tokens are believed to represent the numbers of workmen, such as "one man" or "a gang" (Schmandt-Besserat 1996: 115).

In sum, the scenarios outlined above offer CC hypotheses about the role of tokens and bullae as "certificates of contributions" or "tax receipts" that played an instrumental role within the communal-despotic, redistributive modes of organization. The primary function of these tokens and bullae appears to have been a means of bureaucratic social control over the fulfillment of in-kind payment obligations to a central public authority, such as a temple (or, equivalently, a means of reducing the avoidance of such payment obligations). This

extends the role of tokens and bullae as simple inventory-keeping devices to a means of accounting for the *fulfillment* of in-kind payment obligations to a central public authority. The token system provided a mechanism of checks (verification) that allowed temple officials to detect the non-compliant “taxation units.”

While this chapter maintains that tokens likely served as “certificates of contributions” that were already fulfilled (i.e. tokens represented goods and services already delivered to the temple), in one of her earlier publications Schmandt-Besserat (1982) raised the question whether tokens could have played a role as a means of control over the fulfillment of *future* payment commitments. More specifically, she argued that an increased complexity of the contribution process (e.g. due to population growth and regional expansion) would make it impossible for temple personnel to keep in mind “the great varieties of foods [...] to be supplied by different parties” (Schmandt-Besserat 1982: 876). In such a context, tokens could be “viewed as standing for pledges of food offerings *to be delivered* by individuals and groups” (ibid.: 876; emphases added). Once the actual goods would be delivered, their types and quantities could be compared to the types and quantities pledged, i.e. those represented by the tokens. Thus, according to Schmandt-Besserat’s (1982) earlier hypothesis, tokens stored in public archives would serve as a means of accounting for (or keeping track of) the goods and services owed to a temple, as well as a means of verification for the goods (and services) delivered. “In this light, Schmandt-Besserat (1982) concluded, “the pristine function of the tokens would be an instrument of control, and, therefore, a germ of power over food production in the hands of an emerging central authority” (ibid.: 876).

Despite this slight difference in emphasis, there is a significant common ground between Schmandt-Besserat’s (1982) position and the arguments presented here: both views emphasize the role of tokens as instruments of bureaucratic control over the fulfillment of in-kind payment obligations to a central public authority. The primary difference is that Schmandt-Besserat (1982, 1996) sees tokens as representations of future payment commitments, while this chapter describes them as indicators of past (i.e. fulfilled) contributions.

It is unfortunate that although the link between tokens, bullae and taxation was never abandoned by Schmandt-Besserat, she did not fully pursue her 1982 research program that seriously questioned the role of tokens and bullae as simple inventory-keeping devices and suggested their instrumental role in the context of in-kind payments of “taxes” and tribute to the ancient Mesopotamian temples. This chapter points to the importance of Schmandt-Besserat’s 1982 research agenda for the development of modern theories of money and its origins.

## **Towards an institutional definition of money**

What definition of “money” should a monetary historian or anthropologist adopt when working on the issues of money’s origin? As a rule, a scholar projects his

favorite definition of *modern* money into ancient history, based upon the theoretical perspective on money he or she adheres to (Laum 1924; Einzig [1948] 1951). In this manner, a Metallist would maintain that the earliest money was a commonly accepted medium of exchange (commodity) in a market context, while a Chartalist would argue that the earliest money was a state-mandated unit of account in which *monetary* taxes were denominated and had to be paid. Note that these alternative definitions of the earliest money are based upon modern institutions such as a market economy (in the case of Metallism) and monetary taxation (in the case of Chartalism). However, to project such definitions of money into ancient history is to imply that modern and ancient institutions and modes of socio-economic organization are equivalent or very similar (Laum 1924; Einzig [1948] 1951).

Note that whereas the primary definitions of money adopted by the Metallist and Chartalist perspectives are clearly distinct, what they share in common is an appeal to a certain mode of allocation in a society. Within the exchange-based framework, allocation is performed via a market exchange of commodities among individuals. Within the Chartalist perspective, allocation is achieved via a state-imposed monetary taxation. In both frameworks, however, money is viewed as a key element in carrying out the chosen mode of allocation in a society. While in the Metallist account money performs a key role as a facilitator of market commodity exchange, in the Chartalist perspective money is viewed as a key element (a unit of account) facilitating a centralized collection of taxes by a central public authority.

Aiming at a CC definition of money, this chapter suggests discarding a *specific* definition of money focused on a *specific* monetary function(s) carried out within a *specific* mode of societal allocation. Instead, institutional economists applying Veblen's CC must adopt a more general definition that would recognize that money is closely linked to the particular modes of transactions and forms of socio-economic integration of different societies (Dalton 1971). In particular, I propose to define money in the following way:

Money is a material object and/or an immaterial concept<sup>10</sup> that performs a key role in carrying out the prevailing mode of allocation in a society. In this, money is universally accepted and/or mandated regardless of its intrinsic value.

This broad definition of money may be easily applied to the ancient Mesopotamian societies where money (in the form of clay- tokens and bullae) performed an instrumental role in carrying out the prevailing communal-despotic, redistributive modes of socio-economic integration. Likewise, this general definition may be easily applied to "modern money" based on the key role(s) that it performs under the modern modes of allocation (such as market exchange or monetary taxation).

## Conclusions

This chapter has provided a critical appraisal of the Chartalist perspective on money's origin. It was argued that the traditional Chartalist monetization framework cannot provide an adequate explanation for the emergence of state money in ancient societies, such as Mesopotamia. The goal of this chapter was to reformulate the Chartalist approach to the origins of money to make it fit with the specific transactional modes and forms of socio-economic integration of ancient Mesopotamia in the fourth millennium BC. In particular, it was argued that money was introduced by the ancient Mesopotamian temples as a means of bureaucratic control over the fulfillment of in-kind payment obligations of the local population. Such payment obligations (i.e. a debt relationship of a population to a temple) were already in place before monetization was introduced. Temple money, or tokens and bullae, serving as "certificates" of fulfilled contributions to a temple, provided a system of checks that allowed temple officials to identify the non-compliant "taxation units." It was shown that money came into being via a process of cumulative causation, emerging as a practical solution to a material problem of maintaining control over fulfillment of in-kind payment obligations to a temple (or, likewise, reducing their avoidance).

Note that this CC account of monetization preserves the fundamental Chartalist arguments, namely that (1) money is introduced (invented) by a central public authority; (2) money is universally demanded, accepted, and mandated regardless of its *intrinsic* value; (3) money is introduced in the context of a debt relationship between a population and a central public authority. At the same time, however, this CC account of money's origin: (1) does not view monetization as a means of inducing a local population into a debt relationship with a central public authority; (2) does not rely upon *monetary* taxation; (3) does not view money as a unit of account in which monetary taxes are denominated and have to be paid. Rather, this CC account relies upon in-kind taxation, and views the earliest money as a certificate of a fulfilled in-kind contribution to a temple. Overall, this chapter embeds the Chartalist perspective on money's origin into specific transactional modes and forms of socio-economic organization of ancient Mesopotamian societies. Thereby the Chartalist theory of money's origin is improved.

## Notes

- 1 In ancient Mesopotamia writing did not take off until c.3100–3000 BC (Schmandt-Besserat 1996: 125).
- 2 For convenience and consistency of exposition, the term "state" will be predominantly used hereafter in reference to a central public authority. However, this is not to imply that the Chartalist account is limited to an introduction of money by a *state*. Rather, it is any sovereign, central public authority that the Chartalist perspective refers to in a context of monetization. Nevertheless, in its *monetary* taxation mechanism, the Chartalist public authority is very similar to a modern state.
- 3 However, in Wray's "hypothetical governor" account of money's introduction (Wray 1998: 54–5), even the threat of violence fails to generate a supply of goods and labor services from the local population.

- 4 For a distinction between private property and possession regimes, and its implications for money use and functions, see Heinsohn and Steiger (2000: 67–73, 81–2, 93–5).
- 5 A unit counted could be either a discrete entity, like a sheep from a flock, or a specific measure of a certain product, such as a bag of grain (Nissen *et al.* 1993: 11).
- 6 The numbers of tokens enclosed into clay bullae varied from two to fifteen. The quantities of goods represented by the tokens enclosed were, on average, an equivalent of about five bushels of grain or five sila of oil (1 sila = 0.82 liters). Most of the tokens enclosed were “plain,” i.e. they represented agricultural staples (Schmandt-Besserat 1996: 46, 49, 103).
- 7 As is well documented, the means of personal identification (namely personal seals) were a privilege of priest-kingship, priesthood, and other temple bureaucracy (Nissen 1988: 77, 79, 117).
- 8 See Laum (1924), Desmonde (1962), and Seaford (2004) for the origins of money in ancient religious practices and rituals.
- 9 1 sila is the equivalent of 0.82 liters.
- 10 See Henry 2004.

## References

- Dalton, G. (1971) *Economic Anthropology and Development: Essays on Tribal and Peasant Economies*, New York: Basic Books.
- Desmonde, W. H. (1962) *Magic, Myth, and Money: The Origins of money in Religious Ritual*, New York: The Free Press of Glencoe.
- Einzig, P. [1948] (1951) *Primitive Money in its Ethnological, Historical and Economic Aspects*, London: Eyre & Spottiswoode.
- Finley, M. I. (1965) *The World of Odysseus*, New York: The Viking Press.
- Goodhart, Ch. ([1998] 2003) “The Two Concepts of Money: Implications for the Analysis of Optimal Currency Areas,” in S. Bell and J. Nell (eds) *The State, the Market and the Euro: Chartalism versus Metallism in the Theory of Money*, Cheltenham, Northampton, MA: Edward Elgar.
- Grierson, P. (1975) *Numismatics*, London: Oxford University Press.
- (1977) *The Origins of Money*, London: The Athlone Press, University of London.
- Heinsohn, G. and Steiger, O. (2000) “The Property Theory of Interest and Money,” in J. Smithin (ed.) *What is Money?*, London, New York: Routledge.
- Henry, J. F. (2004) “The Social Origins of Money: The Case of Egypt,” in L. R. Wray (ed.) *Credit and State Theories of Money*, Cheltenham, and Northampton, MA: Edward Elgar.
- Jasim, S. A. and Oates, J. (1986) “Early Tokens and Tablets in Mesopotamia: New Information From Tell Abada and Tell Brak,” *World Archeology*, 17, 3: 348–62.
- Knapp, G. F. [1905] (1924) *The State Theory of Money*, London: Macmillan.
- Kraay, C. M. (1964) “Hoards, Small Change and the Origin of Coinage,” *The Journal of Hellenistic Studies*, 84: 76–91.
- Laum, B. (1924) *Heiliges Geld: Eine Historische Untersuchung über den Sakralen Ursprung des Geldes*, Tübingen: Verlag von J.C.B.Mohr.
- Menger, C. (1892) “On the Origins of Money,” *Economic Journal*, 2, 5: 239–55.
- Modelski, G. (1997) “Early World Cities: Extending the Census to the Fourth Millennium,” Paper Prepared for the World System Historical Data Group ISA convention, Toronto (March 1997). Available on [www.etext.org/Politics/World.Systems/papers/modelski/geocit.htm](http://www.etext.org/Politics/World.Systems/papers/modelski/geocit.htm).



- Nilhamn, B. (2002) "Tokens of Identity? Small Clay Objects in Near Eastern Archeology," Swedish Seminar Papers in Archeology, 2000–2003, Uppsala.
- Nissen, H. J. (1988) *The Early History of the Ancient Near East, 9000–2000 BC*, trans. E. Lutzeler and K. J. Northcott, Chicago, IL: The University of Chicago Press.
- Nissen, H. J., Damerow, P. and Englund, R. K. (1993) *Archaic Book-keeping: Early Writing and Techniques of Economic Administration in the Ancient Near East*, Chicago, IL: The University of Chicago Press.
- Oppenheim, A. L. (1956) "A Bird's Eye View of Mesopotamian Economic History," in K. Polanyi, C. Arensberg and H. Pearson (eds) *Trade and Markets in the Early Empires*, Glencoe, IL: The Free Press.
- Polanyi, K. [1944] (1962) *The Great Transformation: The Political and Economic Origins of Our Time*, Boston, MA: Beacon Press.
- (1957a) "Aristotle Discovers the Economy," in K. Polanyi, C. Arensberg and H. Pearson (eds) *Trade and Markets in the Early Empires*, Glencoe, IL: The Free Press.
- (1957b) "The Economy as Instituted Process," in K. Polanyi, C. Arensberg and H. Pearson (eds) *Trade and Markets in the Early Empires*, Glencoe, IL: The Free Press.
- Schmandt-Besserat, D. (1981) "Decipherment of the Earliest Tablets," *Science*, 211, 4479: 283–5.
- (1982) "The Emergence of Recording," *American Anthropologist*, 84, 4: 871–8.
- (1996) *How Writing Came About*, Austin: University of Texas Press.
- Schmoller, G. von ([1949] 1963) "Political Economy and its Method," in K. W. Kapp and L. L. Kapp (eds) *History of Economic Thought: A Book of Readings*, New York: Barnes and Noble.
- Seaford, R. (2004) *Money and the Early Greek Mind: Homer, Philosophy, Tragedy*, Cambridge: Cambridge University Press.
- Semenova, A. (2008) "Would You Barter With God? Why Holy Debts and Not Profane Markets Created Money," Working Paper.
- Tymoigne, E. and Wray, L. R. (2005) "Money: An Alternative Story," C-FEPS Working Paper No. 45. Available on [www.cfeps.org/pubs/wp-pdf/WP45-Tymoigne-Wray.pdf](http://www.cfeps.org/pubs/wp-pdf/WP45-Tymoigne-Wray.pdf).
- Veblen, T. (1898) "Why Is Economics Not an Evolutionary Science?," *The Quarterly Journal of Economics*, 12, 4: 373–97.
- (1900) "The Preconceptions of Economic Science," *The Quarterly Journal of Economics*, 14, 2: 240–69.
- (1901) "Gustav Schmoller's Economics," *The Quarterly Journal of Economics*, 16, 1: 69–93.
- (1909) "The Limitations of Marginal Utility," *The Journal of Political Economy*, 17, 9: 620–36.
- Wray, L. R. (1998) *Understanding Modern Money: The Key to Full Employment and Price Stability*, Cheltenham, and Northampton, MA: Edward Elgar.

# Index

- accumulation 27, 92, 93, 95, 96, 99, 100, 158, 159, 161–4  
agency 125, 127, 128  
algorithms, genetic 6  
analogy 6, 7, 9, 15, 107  
anthropology, substantivist school of 179, 182  
area, functional economic 29, 30  
  
biological system 5, 8, 131, 142  
biology 4, 5, 175  
bullae, mesopotamian 184–9, 191  
business 12, 44, 45, 61, 140, 155  
business cycle 2, 78, 80, 164–6  
  
capital 13, 14, 16, 17, 24, 27, 34, 36, 44–8, 50, 52, 53, 56, 57, 61, 68–70, 72–4, 80, 81, 85, 88, 96–8, 102, 144, 156, 158–62, 164–8  
capitalism 27, 93, 95, 96, 99, 155, 158, 165, 166, 170  
capitalists 14, 34, 159, 165, 166  
causation, circular 92, 95, 106–8, 112–14, 125, 132, 134, 139; circular cumulative 3–7, 17, 31, 65, 66, 68, 72, 74, 98, 106, 119, 128, 142, 154, 155, 157, 161–3, 166, 170; cumulative 18, 24, 31, 33, 66, 67, 69, 77–85, 92, 95, 111, 113, 134, 148, 156, 172, 174–8, 181, 186, 189, 191, 192; reciprocal 107, 109, 112, 113; self-reinforcing 1, 3–6, 106, 109, 110, 112, 136, 141; social 7, 107, 108; undirectional 78, 86  
CC, theory of 43, 44–7, 106, 111–13, 130, 144  
CCC 1–4, 8, 9, 49, 51, 56, 91–8, 101, 103, 107–15, 119, 120, 124–8, 130–4, 137–9, 142, 145, 146, 148, 149  
change 2, 4, 7, 18, 25, 58, 78, 80, 81, 86–8, 92, 96, 97, 106, 108, 111, 125–7, 132, 133, 136, 154, 161, 163, 164, 166–8, 170; cumulative 6, 103, 112, 175; endogenous 18, 44–7, 74, 80; evolutionary 176; immaterial 174; material 174; structural 161, 163; technological 2, 16, 26, 28, 29, 45–7, 73, 82, 155, 156, 158, 163–6, 168, 169  
chartalism 178, 179, 181, 183, 191  
circle: vicious 106, 107, 110, 114, 139; virtuous 107  
classical theory 154, 161, 162  
complexity 7, 103, 109, 126, 130–5, 142, 143, 146, 170  
continuity 172–6  
contradictions 92, 96, 98, 99–101, 103  
contributions, in-kind 180, 181, 183, 185–7, 189, 190, 192  
costs 23, 48, 55–7, 72, 80, 82, 83, 87, 96, 130, 133, 134, 138–44, 146, 149, 165; fixed 31, 32; higher 86, 159; lower 15, 47, 48, 61, 83, 160, 166; social 48, 106, 111; sunk 16; transport 31–3, 36, 162  
crisis, societal 107, 139; socio-political 93  
cycle 55, 95, 102, 163; business 2, 78, 80, 164–6; trade 78, 80, 112, 164; vicious 79, 82; virtuous 56, 78, 82, 85  
  
Darwin, Charles 5, 6  
Darwinism, Universal 4, 6  
demand 16, 24, 26, 31, 34, 46–9, 51–6, 58, 61, 65, 66–74, 79–87, 91, 92, 94, 98–100, 102, 103, 140, 141, 156, 158–70  
deterministic 15, 23, 24, 25, 33, 173  
development 6, 13, 16–20, 22, 23, 31, 33, 47, 48, 51, 54, 56, 58, 66, 67, 69, 70, 72–4, 93–6, 97, 99, 101, 103, 110, 130, 133–6, 139, 144–6, 149, 155, 156, 158,

development *continued*

- 160, 163; Australian wine industry 67, 72; industrial 4, 50, 72; institutional 4; Northeast Asian 43, 44, 50, 58, 96; regional 67, 79; stages of 66–9, 72, 74; strategies of 43, 44, 50, 51, 56, 58; technological 56, 72–4
- development process 20, 33, 50, 52, 58, 65, 70, 113, 114, 149
- development state 43, 51, 57, 59, 60
- Dewey, John 7, 109, 120
- dialectics 107, 173
- distribution 14, 24, 26, 30, 33, 36, 57, 84, 85, 110, 114, 142, 166, 168
- dynamics 3, 5, 7, 9, 21, 23, 28, 29, 44, 47, 60, 91, 112, 132, 156, 159, 166, 168, 169; CCC 96, 97, 99; complex 132; cumulative 92; socio-economic 108, 112, 132; system 3, 108–10, 125
- ecological effects 131, 134, 141, 143
- ecological system 109, 119, 131–9, 142, 145, 148
- ecology 131, 148
- economics: development 2, 6; ecological 2, 3, 130–4, 145, 146, 148, 149; equilibrium 1, 3, 17, 60, 113; evolutionary 1–4, 7–9, 91, 113; institutional 1–4, 8, 9, 91, 106, 107, 110, 114, 133, 142, 145; neoclassical 9, 22, 44–7, 77, 81, 162, 174; non-equilibrium 6, 92; macro- 163; political 111, 114, 145; Post Keynesian (*see* Post Keynesian); social 94, 110; substantive 109, 145
- economies, agglomeration 31, 32; capitalist 19, 161; dynamic 45, 46; endogenous 32
- economies of scale 13, 15, 16, 18, 19, 20, 24, 27, 31, 45, 83, 92–4, 99, 100, 141, 170; technological 58–60, 66, 67, 71, 72; external 44, 46–8, 51, 54, 57, 95, 155, 156
- economies of scope 99, 100
- economy: classical political 162, 166; political 91, 92, 94, 95, 101–3, 145, 149, 173
- effects: backwash 97; cumulative 112, 114, 134, 137, 138, 168; ecological 131, 134, 141, 143; Keynesian multiplier 83, 118; spread 95, 97
- endogenous 4, 32, 44–7, 74, 80, 86, 92, 96, 98–100, 102, 113, 132, 163–7

- enterprise 54, 55, 59, 66, 68, 73, 74, 83, 96, 138–41; capitalist 102
- equilibrium 16–18, 20, 22, 34, 49, 78, 81, 95–8, 107, 113, 132, 162; general 2, 8, 9, 95, 100, 120; non- 6, 92, 131
- equilibrium theory 17, 19, 20, 21, 32, 33, 35, 77, 80, 81, 166
- ergodic 24, 25; non- 24, 25, 33
- evolution 126, 163, 174–6
- evolutionary economics 1–4, 7–9, 91, 113
- evolutionary philosophy 174, 176, 177
- evolutionary theory 176
- externalities 15, 30, 35, 94–6, 99
- facts, stylized 19, 83, 94, 95
- feedback 3, 92, 96, 110, 134, 155, 168
- financial instability hypothesis 3, 102
- fisheries 130–2, 134, 136–40, 144, 145, 147–9
- function: allocative 19, 35; production 14, 20, 25–9, 34–6, 45, 47
- game theory 3, 6
- geography, economic 1–3, 13, 30–3, 34, 36
- Georgescu-Roegen, Nicholas 134, 141, 144, 145, 149, 174
- German Historical School 182
- governance, democratic 144
- government 43, 48, 50, 53, 55, 65, 77, 79, 81–4, 88, 95, 98, 99, 113, 121, 126, 127, 147, 148
- growth 13, 16, 17, 19, 20, 24, 29, 30, 33, 34, 44–9, 50, 51, 53, 54, 56, 58–61, 65, 66, 72, 74, 79, 81–8, 124, 130, 135–7, 139, 141, 144, 147, 155, 156, 158, 159, 161, 163–70; demand 46, 51, 82, 83; export-led 49, 52, 83; process of 30, 33, 43, 44, 81, 82, 103; productivity 24, 27, 46, 49, 79, 82, 83, 85–8; stages of 65, 67, 68
- growth model 16, 34, 46, 159
- Hegel, Georg W.F. 107, 173
- history 16, 19, 20, 24, 33, 56, 60, 68, 69, 74, 77, 80, 103, 112, 163, 175, 182, 184, 186, 191, 194
- hypothesis 132–4, 142, 173, 176; certificate of contribution 186–90, 192; financial instability 3, 102; Prebisch-Singer 103; research 106, 108, 110
- indicators 147; social 119; socio-economic 133; substantive 142

- industrialization 66–8, 114, 130, 136
- industry 156, 158, 160, 166–8
- inequality 36, 94, 94–7, 110, 113, 139, 143
- innovations 73, 94, 99, 103, 160, 166, 170
- instability: financial 3; macro-economic 3
- institutional economics 1–4, 8, 9, 91, 106, 107, 110, 114, 133, 142, 145
- institutionalism 102, 106, 111
- institutions 96, 102, 106, 108, 110, 114, 119, 125–7, 131, 148, 159
- instrumentalism 7, 119
- interaction 31, 36, 45, 67, 68, 72, 73, 91, 92, 94, 95, 97, 103, 107–9, 131, 138, 139, 142, 148, 167, 176; cumulative 109; reciprocal 107–9, 134, 135
- interdependence 48, 49, 51, 73, 92, 93, 96, 97, 99, 108, 114, 132, 142, 145, 157, 167, 168, 170
- interplay 174–6
- interrelatedness 16, 35, 107, 114
- interrelation 109, 114, 119, 132–4, 139, 142, 148
- investment 16, 24, 34, 35, 44, 46–9, 51, 53, 54–7, 59, 60, 67–70, 78, 83, 85, 87, 93, 99, 100, 103, 112, 159, 163
- Kaldor, Nicholas 1, 3, 4, 8, 13–21, 24, 25, 27, 28, 31, 32–6, 43–5, 47, 49, 50, 58, 61, 65–8, 72, 74, 77, 79–87, 88, 91–4, 101, 103, 106, 113–15, 144, 154–6, 160
- Kapp, K. William 1–3, 6–8, 106–11, 113–15, 130, 133, 134, 136, 137, 139–45, 148, 149
- Keynes, John M. 3, 78, 85, 101, 165, 168, 170
- Keynesian multiplier effect 83, 118
- Knapp, Georg F. 178, 179
- knowledge 22, 33, 50, 55, 59, 73, 92, 94, 97, 98, 119, 127, 131, 133, 134, 142, 146, 148, 157, 172–4
- labor 12, 14, 15, 25–7, 46, 48, 31, 34, 36, 52, 54, 55, 61, 80, 88, 96, 99, 101–3, 156, 158, 159, 163–5, 168, 169; productivity of 52, 83, 86, 162
- Laum, Bernhard 179, 191
- learning 18, 45, 47–50, 53, 59, 99, 148, 155
- lock-in 95, 99, 100, 112
- manufacturing 45, 48–54, 59, 65, 66, 68, 74, 91, 94, 100, 114, 130, 144, 155, 158, 161
- market 15, 17, 19, 20, 25, 31, 33, 35, 49, 50, 53, 55, 57–61, 66–74, 77, 78, 80, 82, 102, 130, 141, 142, 144, 147, 156, 158–60, 162–4, 166, 170; failure of 22, 51, 60
- Marshall, Alfred 15, 21, 34
- Marx, Karl 13, 14, 26, 107, 110, 140, 145, 154, 155, 158, 160, 162–6
- Marx's theory 164–6
- matrix, social fabric 119–22, 125, 127, 128
- Menger, Karl 179
- mesopotamia, ancient 184–7, 189, 190
- mesopotamian bullae 184–9, 191
- metallism 179, 183, 191
- metaphor 2, 6, 7
- method 77, 93, 94, 142, 143, 146, 165, 166, 172, 175, 176
- methodology 17–20, 32, 33, 36, 43, 46, 61, 65–7, 74, 80, 108, 109, 131, 133, 134
- minima, social 110, 111, 131, 144–7
- Minsky, Hyman 3
- Model: CCC 101; cumulative causation 18, 24, 33, 60; four-stage 50, 66, 74; growth 16, 34, 46, 159; Kaldor's 31, 67, 93; Krugman-Venables 31; NEG 31–3, 36; neoclassical 14, 22, 24, 34, 35, 162; Pasinetti's 168, 170; Solow-Swan 16, 24; structural 167–70
- monetization 178–81, 183, 184, 186, 189, 192
- money 78, 85–7, 92, 93, 96, 102; origin of 178–82, 186, 190, 192; as token 179, 180, 184
- motion, CCC 98
- Myrdal, Gunnar 1–9, 16, 17, 31, 49, 65, 66, 74, 78–81, 91–4, 96, 97, 101, 103, 106–15, 119, 125, 130, 132, 133, 139, 142, 143, 148, 149, 154
- neoclassical economics 9, 22, 44–7, 77, 81, 162, 174
- networks 96–8, 101, 103, 119, 127, 140, 148
- New Economic Geography *see* geography, economic
- non-equilibrium 6, 92, 131
- non-ergodic 24, 25, 33
- non-teleological 5, 9, 113
- open system 4, 102, 108, 109, 111, 120, 124
- openness 3, 4
- pathdependence 3, 15–17, 21, 22, 24, 32, 33, 35, 92, 95, 97, 102, 112
- Peirce, Charles S. 172–6, 177

- philosophy, evolutionary 174, 176, 177  
 planning 43, 48, 54, 57, 80, 108, 109, 111, 130, 167  
 Polanyi, Karl 109, 145, 179, 182, 188  
 policy 4, 5, 7, 31, 52, 65, 77, 78, 81, 83–8, 95, 113, 119, 140, 144, 167;  
   development 49, 61; economic 2, 44, 83, 88, 145; government 4, 66, 97; industry 44, 50–2, 54–9, 61, 81, 83  
 policymaking 4, 111, 114  
 political economics 111, 114, 145  
 political economy 91, 92, 94, 95, 101–3, 145, 149, 173  
 pollution 111, 130, 139, 141, 142, 144  
 Post Keynesian 3, 77, 81, 84, 85, 87, 88, 102, 154, 155, 166  
 poverty 52, 79, 96, 97, 103, 106, 107, 110, 111, 114, 139, 149  
 process 23, 24, 27, 33, 35, 45, 53, 59, 66, 67, 70, 73, 78, 92–6, 98, 100, 102, 103, 106, 108–13, 119–21, 126, 127, 131–4, 139–41, 145–8, 154, 156–60, 166–70, 172, 174–6; CCC 93, 95, 65–7, 74;  
   circular 102, 103; cumulative 58, 60, 65, 79, 80, 83, 85, 86, 88, 107, 108, 112, 133, 166, 175; development 20, 33, 50, 52, 58, 65, 70, 113, 114, 149; economic 20, 78, 93, 97, 108, 113, 174; non-ergodic 25; non-linear Polia 35; positive feedback 92; random 27; revolutionary 175; stochastic 21, 23–5  
 production 13–15, 17–21, 25–7, 32, 35, 44–8, 50, 52–6, 58, 59, 61, 66, 69, 71–3, 79, 82, 85, 86, 102, 108, 135, 136, 139–42, 147, 155, 156, 158–70  
 production function 14, 20, 25–9, 34–6, 45, 47  
 productivity 12–14, 16, 17, 24, 26, 27, 35, 36, 45–52, 54, 56, 59, 79, 82, 83, 85–8, 155, 156, 159, 160, 162, 165, 169  
 profit 13, 14, 20, 32, 47, 55, 85, 87, 96, 102, 141, 142, 146, 148, 159–66  
 reciprocal causation 107, 109, 112, 113  
 reproduction 101–3, 110, 144, 148, 149, 164, 167  
 resources 20, 21, 23, 44, 45, 47, 53, 74, 79–82, 85, 97, 98, 101, 102, 131, 132, 135–41, 144–9, 161, 163  
 returns, constant 13, 14, 16, 25–9, 35, 44; diminishing 14, 34, 44, 47, 140, 144, 161, 162; ecological 144; increasing 13–16, 18, 19, 21, 24, 25–32, 34–6, 44–7, 48, 50, 51, 53, 54, 57–9, 65, 79–83, 113, 114, 130, 144, 149, 155, 156, 161, 162; pecuniary 141, 144  
 Ricardo 13, 14, 16, 155, 160–6  
 salmon 130, 131, 134–41, 144, 147–9  
 Schmoller, Gustav von 182  
 Schumpeter 101, 103, 126, 127  
 science 172–6  
 self-organization 3, 6  
 self-reinforcing causation 1, 3–6, 106, 109, 110, 112, 136, 141  
 Smith, Adam 12–17, 25, 79, 81, 82, 91, 154–61, 164–6  
 social costs theory 130, 133, 141, 142  
 social fabric matrix 119–22, 125, 127, 128  
 social indicators 119  
 social system 7, 96, 108, 111, 112, 114, 119, 134  
 socio-economic dynamics 108, 112, 132  
 specialization 12, 13, 17, 20, 45, 66, 94, 102, 103, 139, 156, 160  
 stages 15, 34, 66–70, 72, 74, 112  
 stages of growth 65, 67, 68  
 state 4–6, 113; development 43, 51, 57, 59, 60  
 stylized facts 19, 83, 94, 95  
 substantive economics 109, 145  
 substantive indicators 142  
 substantivist school of anthropology 179, 182  
 supply 31, 46–9, 57, 70, 81–3, 86, 87, 91, 93, 94, 98, 99–101, 103, 137, 145, 148, 159–61, 163–5  
 surplus value 102  
 synechism 173, 174, 176  
 system, biological 5, 8, 131, 142; closed 143; complex 109, 125; ecological 109, 119, 131–9, 142, 145, 148; economic 77, 107–9, 111, 131, 134, 142, 145, 154, 156, 163, 166, 170; open 4, 102, 108, 109, 111, 120, 124; social 7, 96, 108, 111, 112, 114, 119, 134; tax 84, 85  
 system dynamics 3, 108–10, 125  
 tax 78, 81, 83–5, 87, 88, 123, 145  
 technological change 2, 16, 26, 28, 29, 45–7, 73, 82, 155, 156, 158, 163–6, 168, 169  
 technological development 56, 72–4  
 technological economies of scale 58–60, 66, 67, 71, 72  
 technology 119, 131, 139, 141, 144, 148, 165, 167  
 teleology 4, 108, 112, 176; non- 5, 9, 113

- tendencies 5, 18, 60, 95, 97, 112, 134, 140, 160, 165, 173, 176
- theory, complexity 109, 142; classical 154, 161, 162; development 101; economic 95, 132, 163; equilibrium 17, 19, 20, 21, 32, 33, 35, 77, 80, 81, 166; evolutionary 176; game 3, 6; growth 13, 16, 30, 31, 34, 50, 82; Marx's 164–6; neoclassical 17, 20; open system 134; social costs 130, 133, 141, 142; systems 109, 120, 134; trade 13, 30, 36; value 14, 15, 154
- thermodynamics 8, 134
- trade 49, 52, 56, 67, 78, 80, 99, 156, 163
- trade cycle 78, 80, 112, 164
- trend 4–7, 18, 103, 107, 112, 148, 165
- uncertainty 25, 32, 87, 99, 100, 102, 132, 139, 142, 143, 145–7
- values 5–7, 51, 93, 94, 110, 114, 130, 140, 143, 163, 164; surplus 102
- Veblen, Thorstein B. 1–5, 8, 15, 16, 91, 101, 106, 111, 113, 115, 132, 133, 140–2, 145, 154, 172–8, 181, 182, 186, 191
- Verdoorn's Law 27–9, 31, 36, 49, 61
- vicious circle 79, 82, 106, 107, 110, 114, 139
- virtuous circle 56, 78, 82, 85, 107
- wealth 6, 87, 96, 97, 130, 140, 149, 160
- welfare 13, 21, 56, 120
- Young, Allyn 15–18, 34, 45, 47, 60, 74, 79, 80, 82, 91, 154, 156–9